This electronic thesis or dissertation has been downloaded from the King's Research Portal at https://kclpure.kcl.ac.uk/portal/

Interpreting Gödel
Historical and Philosophical Perspectives

Chen, Long

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. May. 2023

Interpreting Gödel: Historical and Philosophical Perspectives

Long Chen

A Thesis Submitted for the Degree of

Doctor of Philosophy

September 2017

Abstract

The main project of the thesis is to provide a comprehensive examination of Gödel's philosophy of mathematics and logic, in defence of his Platonism and the idea of mathematical intuition. The whole work is considered to be partly historical, explaining Gödel's ideas and arguments in comparison with his contemporaries such as Carnap, Russell, Turing and Hilbert; partly mathematical, relying especially on his incompleteness theorem to refute certain philosophical positions; and partly philosophical, shedding light on debates in current discussions of philosophy of mathematics and logic such as the nature of paradoxes, the indispensability argument, the feasibility of formalism, etc.

After the introduction, chapter one deals with Gödel's critique and refutation of Carnap's syntactical view of mathematics, which interprets mathematics purely in terms of linguistic rules of symbols, and thus void of content. Chapter two discusses Gödel's criticism of Russell's constructivistic view towards logic, especially his no-class theory and the vicious circle principle. Chapter three turns attention to Gödel's role in the development of computability theory, using it as a case study for conceptual analysis and tries to reconcile the apparent conflict concerning Gödel's remarks on Turing, who gave a "precise and adequate definition" of mechanical computability and yet committed "a philosophical error" in his argument. The last chapter focuses on Gödel's relation to Hilbert, especially the relation of his incompleteness theorems for Hilbert's finitary consistency proof program and the significance of Gödel's serious

engagement with finitism. The thread uniting all the discussions in the chapters is the idea of the incompletability or inexhaustibility of mathematics by pure formalism and the indispensability of abstract concepts and a kind of mathematical intuition in providing a satisfactory foundation for mathematics and in enabling the human mind to surpass any particular mechanism, which Gödel himself considered to be the major philosophical conclusion to be drawn from his logical results.

Acknowledgements

I want to thank, first of all, my parents for their continual support and encouragement, allowing me to choose whatever I wish to do and wherever I like to do it.

The biggest support during my whole PhD came from my main supervisor, Michael Beaney, whose seminars back in Peking University aroused my interest in the philosophy of mathematics and logic and who gave me lots of suggestions and advice for writing this thesis and also plenty of help in adapting myself to life in the UK. My former supervisor Mary Leng was always very nice and helpful and I learned a lot from our discussions and her feedback on drafts of the early chapters. Wilfried Meyer-Viol greatly enhanced my knowledge of logic in our fortnightly meetings in his private time, which benefited me a lot. I am indebted to Nils Kürbis and Julien Dutant in conversations about logical and philosophical problems. The London Philosophy of Mathematics Reading Group was also very enjoyable and stimulating; lots of interesting discussions happened with Marcus Giaquinto, Neil Barton, Josephine Salverda, Clare Moriarty, John Heron and Alex Franklin.

Among my friends and fellow PhD students, I want to thank first Daniel Molto, David Price, Suki Finn and Bob Clark from York, who were very friendly and helpful

in philosophical discussions. In particular, I need to mention a group of friends at King's College London who have together made the philosophy department a very nice community. To Fintan Mallory, with whom I had lots of interesting discussions about logic and language and who kindly helped to proofread one of my chapters; to Paul Doody for proofreading, lots of helpful suggestions and the funny conversations; to Clare Moriarty for becoming my invincible pool partner; to Jørgen Drystad for teaching me chess and appreciating classical music. And also to Dave Jenkins, Dr Dave Preston, Samuel Kimpton-Nye, Dr Tuomas Pernu, Sérgio Farias, Dr Dimitri Mollo, Nate Oseroff, and Will Sharp. Especially I want to thank my friend Qi Li from the film study department, for giving me invaluable support and help in the toughest days of my PhD.

This thesis is written while I was funded by CSC (Chinese Scholarship Council), to which I have to express my gratitude.

Contents

Introduction

The proof theorist Gaisi Takeuti once wrote "from the perspective of logicians, Gödel seems godlike" (Takeuti 2003, 35). The achievement of Gödel qua logician is indeed hard not to notice, as it penetrates nearly every branch of modern logic. The most felicitous remark about the greatness of Gödel's work, in my view, remains the one given by von Neumann on the occasion of Gödel's receiving the Einstein Award in March 1951:

> Kurt Gödel's achievement[1] in modern logic is singular and monumental – indeed it is more than a monument, it is a landmark which will remain visible far in space and time. Whether anything comparable to it has occurred in the logic of modern times may be debated. In any case, the conceivable proxima are very, very few. The subject of logic has certainly completely changed its nature and possibilities with Gödel's achievement. … No demonstrability within mathematics proper had ever been rigorously established before Gödel. The subject of logic will never again be the same. [quoted from (Takeuti 2003, 36–37)]

As for Gödel qua philosopher, the pendulum seems to have gone a full swing.

[1] von Neumann was mainly referring to, among many important achievements of Gödel, the "two absolutely decisive ones" of Gödel's incompleteness theorem and the consistency proof of the axiom of choice and continuum hypothesis relative to ZF.

Some even claim that Gödel was "a logician per excellence but a philosophical fool" (James 1992, 131). Needless to say, especially with the publication of Gödel's posthumous philosophical remarks[2], frivolous remarks of a similar kind cannot be any further from the truth. It does, however, indicate a general sense of suspicion and qualm about Gödel's views from the philosophical community.[3] The most widely discussed and striking aspect of Gödel's philosophy towards mathematics and logic is undoubtedly his "unadulterated"[4] realism or Platonism about mathematical, logical objects and concepts and the relevant notion of mathematical intuition or perception of those objects and concepts. The following much quoted passages should give us a good glimpse:

> The truth, I believe, is that these concepts [concepts denoted by mathematical terms] form an objective reality of their own, which we cannot create or change, but only perceive and describe. (Gödel 1951, 320)

And Gödel says about the "Platonistic view":

> Thereby I mean the view that mathematics describes a non-sensual reality, which exists independently both of the acts and of the dispositions of the human mind and is only perceived, and probably perceived very

[2] See especially (Wang 1996; Gödel 1995; Engelen and Crocco 2015). It is to be regretted that the last one (also the newest one) contains only a partial selection of Gödel's philosophical remarks and all in German. For, I think, Gödel usually explains his ideas much better and clearer than any of the commentators.
[3] According to Parsons, Gödel's robust Platonism "has scandalized many philosophers" (Parsons 1995, 44).
[4] A phrase used by Russell to describe Gödel, see (Russell 1968, 356).

incompletely, by the human mind. (ibid. 323)

The most famous (or infamous) passage about mathematical intuition goes as follows:

> Despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see any reason why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception, which induces us to build up physical theories and to expect that future sense perceptions will agree with them and, moreover, to believe that a question not decidable now has meaning and may be decided in the future. (Gödel 1964, 268)

These passages do show a very strong conviction of Platonism on Gödel's side. However, despite his own declaration that he "was a conceptual and mathematical realist since about 1925"[5] (Gödel 2003a, 444) and his emphasis on the great role that his realistic (or the objectivist) view played in his logical achievements,[6] nothing in his work (include the unpublished lectures and letters) would suggest the existence of a mature philosophical theory comparable to his logical work in terms of insightfulness,

[5] The credibility of this sentence has been doubted by invoking some seemingly conflicting remarks made by Gödel in the 1930s, see (Parsons 1995; Davis 2005). It is not our main task to provide a historical development of Gödel's philosophy, but I tend to think that that evidence can be explained away without contradicting Gödel's claim; see (van Atten and Kennedy 2009; Xing 2011) for such a possible interpretation.
[6] See particularly Gödel's letter to Hao Wang, in (Wang 1974, 8–10).

precision and completeness. In an approved text by Gödel himself he admitted that "[he] has never arrived at what he looked for: to arrive at a new view of the world, its basic constituents and rules of their composition" (Wang 1981, 659), a view which Plato, Descartes (and maybe Husserl) claimed to have had. Yet these dismaying facts should not prevent someone from extracting a plausible and coherent position in Gödel's philosophy, if we bear in mind what Gödel was looking for from philosophy was of an extremely high standard. In the same place where he describes the Platonistic view, Gödel also expresses the hope[7] that

> [A]fter sufficient clarification of the concepts in question it will be possible to conduct these discussions with mathematical rigor and that the result then will be that (under certain assumptions which can hardly be denied [in particular, the assumption that there exists at all something like mathematical knowledge]) the Platonistic view is the only one tenable. (Gödel 1951, 322)

This particular conception of philosophy, whose discussions can be as rigorous as mathematics, might be the real reason for the scarcity of philosophical discussions and lack of a systematic theory in Gödel's works, for "in view of widely held prejudices, it may do more harm than good to publish half done work" (Gödel 2003b, 244). Gödel might be displaying his usual caution,[8] but this cannot prevent other people from taking his position and attacking or defending it one way or the other. About Gödel's

[7] In his introductory note, Boolos describes this hope as "utterly strange" (Boolos 1995, 303).
[8] For more about Gödel's caution, see (Solomon Feferman 1984).

Platonism, lots of people have questioned, not the validity, but the adequacy of his arguments for the desired position. Michael Potter, for example, has argued that, based on the arguments Gödel actual gave, his conception of mathematical intuition of abstract concepts "bears a striking similarity to Dummett's conception of our grasp of indefinitely extensible concepts" (Potter 2001, 331) and "there is hardly anything I have said so far [Potter's interpretation of Gödel's arguments] with which a Dummettian intuitionist need disagree" (ibid. 343). So, after all, Gödel, at least the one "on paper", was not up to a real "Gödelian Platonist". A similar objection has been raised by Charles Parsons, who pointed out that "the main difficulty … however, is not the omission … of a case against Aristotelian realism and psychologism,[9] but that its central arguments are meant to be independent of one's standpoint in the traditional controversies about foundations" (Parsons 1995, 55). There seems to be a dilemma here for Gödel: he wants to make his philosophical argument (just like his logical proofs) as general as possible, i.e., acceptable from a position different from his own Platonistic view such as intuitionism or even finitism. In doing so, however, his argument will become a common force for all those different positions and lose its unique character for Platonism.[10] A weaker objection has also been raised by William Tait, who distinguishes two senses of Platonism or realism in mathematics, one being "the view that terms in mathematical propositions, such as propositions in number theory or set theory, denote *sui generis* objects, i.e., which are not physical objects ,

[9] In the Gibbs lecture, after he "disproved the nominalistic view" (see the chapter on Carnap and Gödel for more details), Gödel listed Aristotelian realism and psychologism as the only two alternatives to Platonistic realism.
[10] We won't dwell on this problem here at the moment, suffice it to say that this might cause problems for the radical pluralist who believes that we can discuss all philosophical positions on an equal footing. For more about this point, see the discussion of the principle of tolerance in the chapter on Gödel and Carnap.

nor mental objects nor fictions, for example, and are not analyzable away" (Tait 2001, 102), which Tait called the "default view" (ibid.) in contrast to the other more substantial one, the claim that mathematical objects really exist, based on "a convincing and non-question-begging analysis of what it means to exist" (ibid. 103). Gödel's argument, according to Tait, "is simply an argument for Platonism in this default sense" (ibid.) which, however, "does not require for its justification the elimination of alternatives—that it be proved in Gödel's sense" and "is not a substantive philosophical position" (ibid.) and can be trivially true.[11] By a similar token, Donald Martin, after distinguishing two senses of 'concept', one in the structural sense and the other in the straightforward sense as applying to objects, argues that "Gödel is entitled to hold that the general iterative concept of set has (or in a different case the axioms for sets of integers implicitly defined)—at least, up to isomorphism—a concept of set (or set of integers) in the straightforward sense, provided that there is such a concept" (Martin 2005, 217), but, he insists, there still remains the issue of "whether there are any instances and whether it matters that there are" (ibid.).[12] So in the same way that we cannot differentiate Gödel from an intuitionist like Dummett as Potter argues, there is nothing in Gödel's arguments that separates himself from a structuralist as Martin sees it.[13]

[11] This seems to be a return to Carnap's distinction between external, pragmatic existence problems and internal ones; see (Carnap 1950a). This kind of pluralism and the relevant minimalist conception of philosophy we will discuss in the chapter on Gödel and Carnap.

[12] Charles Parsons has expressed a similar view when he says that even if we grant Gödel everything he could wish for concerning intuition of the objects of higher set theory, "it is far from clear that he has a case for the transcendental realism concerning these objects that he seems to adhere to" (Parsons 1995, 71).

[13] In a similar manner, Cassou-Noguès has argued that Gödel's main arguments cannot even separate himself from a constructivist who only requires that the process of creation by the human mind "cannot be completely analyzed and re-enacted", a position much weaker than Gödel's conceptual realism; see (Cassou-Noguès 2005, 211).

As for the closely related notion of mathematical intuition[14], the reception fares little better. Charles Chihara's comments below represent the kind of suspicion not unusual among philosophers:

> Whether or not one finds Gödel's reasoning concerning mathematical objects to be at all plausible, even supporters of the Gödelian view must admit that there are features of the view that make it difficult to accept. The mathematician is pictured as theorizing about objects that do not exist in physical space. This makes it appear that mathematics is a very speculative undertaking, not very different from traditional metaphysics. A mysterious faculty is postulated to explain how we can have knowledge of these objects. Gödel's appeal to mathematical perceptions to justify his belief in sets is strikingly similar to the appeal to mystical experiences that some philosophers have made to justify their belief in God.[15] Mathematics begins to look like a kind of theology. It is not surprising that other approaches to the problem of existence in mathematics have been tried. (Chihara 1990, 21)

So for Chihara, the assumption of the existence of mathematical intuition, which is essentially solipsistic and mystical, not only cannot undertake the epistemological role

[14] Parsons tries to separate the two aspects of Gödel's philosophy, claiming that Gödel's notion of mathematical intuition "is not quite so intrinsically connected with his Platonism as one might think" (Parsons 1995, 45). However, I think his arguments are valid only towards Platonism about objects. Considering Gödel's adherence to conceptual realism, it is hard to see how the notion of intuition can be separated from this realism.

[15] In an earlier paper Chihara also compared the experiences of set-theoretic axioms forcing themselves upon us as being true to the vaguely described "mystical experiences" (Chihara 1982, 215), and he thought both will fail to justify belief in the existence of the desired objects, be it sets or God.

expected, but what's worse, will make mathematics degenerate into "a kind of theology". A much milder objection comes from John Burgess. Burgess doesn't think Gödel's notion of mathematical intuition is a mystical notion, but he denied the necessity for assuming such a kind of intuition. After a classification of non-sensory intuitions into the three main types of (a) rational (mainly mathematical, set-theoretical included), (b) linguistic and (c) heuristic, Burgess tries to answer that the phenomenon Gödel wanted to explain by appeal to (a) could also be explained by appeal to (b) and (c), and his answer is "a tentative 'no'" for (b), and "a tentative 'yes'" (Burgess 2014, 31) for (c). So according to Burgess, the existence of a mathematical intuition is real, but it is much more, so to speak, "mundane" than Gödel conceived, it is after all only a heuristic principle, not an epistemological one. An even more modest objection was raised by Charles Parsons. Parsons agrees with Gödel that we do have a certain amount of intuitive knowledge[16], but this does not go very far, certainly not to the extent that Gödel would wish them to be, for example, in the new axioms of infinity in set theory. In particular Parsons thinks that "the complex, iterated reflection involved in the uncovering of stronger mathematical axioms and the concepts entering into them strikes me intuitively as very different from perception" (Parsons 1995, 64) and would be better called "a theory of reason" (ibid.). On the other hand, there do exist attempts to interpret the full notion of mathematical intuition (including, or most importantly, the set-theoretic one) from a positive point of view. For example, Penelope Maddy has attempted to demystify the notion of mathematical intuition by trying to give a

[16] See (Parsons 1979, 1998) for example. We will discuss the nature and limits of this intuitive knowledge and finitism in more detail in the chapter on Gödel and Hilbert.

naturalistic interpretation of our perception of sets.[17] From a somewhat different direction, i.e., by focusing on the notion of intuition in the phenomenological tradition, especially Husserl to whom Gödel referred to explicitly in his later works[18], Richard Tieszen has presented a number of papers arguing for the legitimacy of an intuition of abstract concepts and a new type of Platonism based on it.[19]

However, it strikes me that there is a huge difference between the way Gödel himself conducted his arguments and how commentators interpret them. Gödel's arguments for Platonism and mathematical intuition are always local rather than global, i.e., always in a certain mathematical or logical context and are usually related to or supported by certain concrete mathematical/logical programs or theorems, which is nothing but a demonstration of his claim mentioned above that Platonism as a working philosophy is extremely important for his logical discoveries. Likewise, my aim in the following chapters will be concerned less with such purely philosophical debates as with understanding and examining Gödel's philosophical arguments in a contextual and historical way, keeping the above objections and possible interpretations in the background with the intention that the detailed examinations will provide a solution or response to some of them and clear possible confusions away. My aim here will in fact be less ambitious and more realistic than my title would suggest. I will be concerned not with anything and everything that Gödel has had to say as a logician and philosopher, but more specifically with his ideas about

[17] See in particular (Maddy 1980) and the relevant chapter in (Maddy 1990).
[18] See especially (Gödel 1961).
[19] See (Tieszen 1984, 2002, 2011) for example. Also see (Atten and Kennedy 2003) for an account of Gödel's philosophy in relation to Husserl.

mathematics and logic in a piecemeal manner,[20] the central point being that mathematics does have a content, in fact an incompletable one that cannot be exhausted by any formalism, but that can be grasped by the human mind with an intuition of its concepts. I will discuss and defend Gödel's philosophical ideas in his exchanges, in particular. with Carnap, Russell, Turing and Hilbert.[21]

In the first chapter I will discuss Gödel's critique of the syntactical point of view of mathematics, represented by the logical positivists and in particular Carnap, focusing on Gödel's paper "Is Mathematics Syntax of Language?" (Gödel 1953/9a, 1953/9b). The refutation of such a linguistic account of mathematics, a view which tries to reduce mathematical content and mathematical intuition by syntactical rules about symbols, will confirm the opposite of it, i.e., that there do exist irreducible mathematical facts and intuitions. The so-called indispensability arguments for mathematical realism and philosophical pluralism will be discussed along the way.

With a similar strategy in mind, I will discuss Gödel's critique of Russell's mathematical logic, especially his constructivistic tendency towards mathematical objects embodied in his no-class theory and the theory of types in the second chapter, mainly by analyzing Gödel's 1944 paper on Russell (Gödel 1944). By showing the difficulty in Russell's logical system to build up mathematics without the assumption of the existence of sets or mathematical/logical concepts we will come to see that a realist position is the more defensible one. The epistemological principle of inductive

[20] Thus I have to omit Gödel's very interesting ideas about the relationship between relativity and Kant's philosophy, his cosmology, his ontological argument, etc., but only mention them when necessary.
[21] This choice is made on the basis of the contextual principle I mentioned above, namely, all of them have something mathematical or logical to exhibit along with their philosophical ideas.

justification for new axioms, the slingshot arguments, the vicious circle principle, and the merit and weakness of the theory of types will be our main topics.

The intensional paradoxes seem to pose a real threat for a conceptual realist and for the possibility of a type-free system for the theory of concepts, an idea which Gödel hints at in the paper on Russell. We will turn our attention, in the third chapter, to the positive account of maybe the most successful example of conceptual analysis, i.e., Turing's analysis of mechanical computability. Turing's achievement, on the one hand, constitutes a paradigm case for Gödel's conception of perceiving an abstract concept and, on the other hand, provides the ultimate justification for the generality of Gödel's incompleteness theorem and makes possible a much more fruitful discussion of the relationship between formalism and intuition, minds and machines. We will discuss the fascinating history of the almost simultaneous occurrences of different logically equivalent characterizations of this concept (one of them turns out to be Gödel's own formulation) and contrast Gödel's hesitation about the others (especially Church's) with his conviction of Turing's analysis for an illuminating exhibition of the intuition (or perception) of concepts. Gödel's discussion of "Turing's philosophical error" (Gödel 1972b) will constitute, however, the other side of the whole story, where Gödel challenges Turing's identification (as understood by Gödel) of mechanical procedures with mental ones. We will try to solve the apparent conflicts of Gödel's remarks on Turing and bring out the really distinctive element in Gödel's criticism of Turing, namely the non-mechanical nature of our intuition and reflection on abstract concepts.

This idea of the necessity of the existence of an abstract intuition for concepts

based on meaning rather than the finite combinatorial properties of symbols for a satisfactory foundation of mathematics is brought to the fore in a most striking way in Gödel's lifelong engagement with Hilbert's finitism and constructive consistency proofs, without which any discussion of Gödel's philosophy of mathematics would be incomplete. In the fourth (and the final) chapter we will first present the much discussed problem about the impact of Gödel's incompleteness theorems on Hilbert's program to give a finitary consistency proof for classical mathematics. We will argue that, given certain generous assumptions, Hilbert's program with its original philosophical aim indeed cannot succeed in the face of Gödel's theorems. Based on Gödel's discussions about Hilbert's program and finitism over the years, we will try to reach a conclusion about his evaluation of them and reconcile the apparent problem of Gödel's Platonism and his serious engagement with finitism and constructive consistency proofs by focusing on the notion of abstract intuition.

1. Gödel and Carnap: Syntax and Intuition

1.1 The Significance of the Discussion Between Carnap and Gödel

In 1995, with the publication of volume 3 of his collected papers (Gödel 1995), mainly unpublished essays and lectures during his lifetime, Gödel's sustained effort during the 1950s to refute the conception of mathematics of the Vienna Circle (represented particularly by Carnap)—which he called 'a combination of nominalism and conventionalism' (Gödel 1953/9a, 334)—was made widely available and his covert hostility towards positivism was finally made explicit.[22] Among these efforts the most important and philosophically mature work is definitely the paper entitled "Is Mathematics Syntax of Language?". This was to have been Gödel's contribution to the Carnap volume in *The Library of Living Philosophers,* a series to which Gödel had contributed twice already.[23] Gödel accepted the invitation for the paper in 1953 and

[22] In a letter reply to Grandjean about his relationship with the "Vienna Circle", Gödel admits that his interest in the foundations of mathematics was aroused by them, especially Carnap's lectures on foundation of mathematics and logic, but he "never held the view that mathematics is syntax of language. Rather this view, understood in any reasonable sense, can be disproved by my results"(Gödel 2003a, 444). Among the unpublished works, Gödel first criticized Carnap in the 1951 Gibbs lecture, where he presented Carnap's syntactical view of mathematics as "the most precise, and at the same time most radical, formulation" (Gödel 1951, 315) of the more general view that mathematics is our own creation, a view which Gödel tries to refute in the lecture using results from foundational researches, especially Turing's and his own logical theorems. The discussion there overlaps considerably with and is much briefer than the main text which we are going to examine in this essay. Another place where Gödel expressed his disagreement with the general positivistic view can be seen from a quotation below about different philosophical world-views. For an overview of Carnap's intellectual relationship with Gödel, see (Goldfarb 2005) and (Awodey and Carus 2010) from a logical and philosophical point of view, respectively.

[23] Namely, the Russell volume (Gödel 1944) and the Einstein volume (Gödel 1949). During his lifetime Gödel was invited by the editor Paul Schilpp to contribute four times for these series. The Einstein volume was the most complete one, with Gödel's essay and Einstein's reply, in Russell's case without Russell's reply due to external reasons and Russell's tacit acceptance to Gödel's criticism. Gödel promised but eventually didn't publish a paper for the volume on Carnap and declined the invitation to contribute a paper for Popper's volume.

actually had finished six versions[24] of different length around 1959 but none of them seemed to him totally satisfactory and in the end he decided not to publish it at all. The main reason Gödel gave for his reluctance to submitting the paper, apart from the fact that it would be unfair to Carnap because it was too late to allow him to frame a reply, is that:

> The fact is that I have completed several different versions, but none of them satisfy me. It is easy to allege very weighty and striking arguments in favor of my views, but a complete elucidation of the situation turned out to be more difficult than I had anticipated, doubtless in consequence of the fact that the subject matter is closely related to, and in part identical with, one of the basic problems of philosophy, namely the question of the objectivity of concepts and their relations. On the other hand, in view of widely held prejudices, it may do more harm than good to publish half done work. (Gödel 1953/9, 324)

Despite Gödel's typical caution and reservations, this rich and remarkable work has stimulated extensive philosophical attention and discussion. Some commentators tend to dismiss Gödel's doubt and are convinced of his arguments. Others are on Carnap's side, trying to point out a fallacy in Gödel's argument or undermine his

[24] These versions differ mainly in ways of formulation and the extent of the details of arguments and side issues, without any substantive variation in the core philosophical parts. Version III and V are published in (Gödel 1995) while II and VI are in (Rodriguez-Consuegra 1995).

arguments against Carnap. My aim in this chapter is to give a comprehensive and critical assessment of Gödel's arguments in his critique of Carnap and the contemporary philosophical commentaries relating to this debate. The significance for this survey is threefold. First, we can gain a better understanding of Gödel's philosophy, especially his Platonism and the notion of mathematical intuition, for "the objective of the syntactical program can … be stated thus: To build up mathematics as a system of sentences valid independently of experience, without using mathematical intuition or referring to any mathematical objects or facts" (ibid. 335). The failure of the syntactical program would, even if not in a conclusive way, definitely lend more plausibility to Gödel's own view that mathematical intuition as well as mathematical objects and facts exist. Secondly, through a careful analysis of Gödel's argument, Carnap's views on the foundation of logic and mathematics in his syntactic stage (represented mainly in *The Logical Syntax of Language* (Carnap 1937), *LSL* hereafter) will become manifest, with all its novelties and inadequacies. The last point may not be obvious, but still very important, if we consider both Gödel's and Carnap's philosophy of mathematics and logic from the general "schema" of a variety of possible different views. Actually, in a lecture draft named "The modern development of the foundations of mathematics in the light of philosophy" around 1961, Gödel set up a general schema of possible philosophical world-views (*Weltanschauungen*) in order to discuss foundational research in terms of philosophical concepts:

> I believe that the most fruitful principle for gaining an overall view of the

> possible world-views will be to divide them up according to the degree and the manner of their affinity to or, respectively, turning away from metaphysics (or religion). In this way we immediately obtain a division into two groups: skepticism, materialism and positivism stand on one side, spiritualism, idealism and theology on the other. We also at once see degrees of difference in this sequence, in that skepticism stands even farther away from theology than does materialism, while on the other hand idealism, e.g., in its pantheistic form, is a weakened form of theology in the proper sense. (Gödel 1961, 375)

Gödel clearly identifies himself with the "right" by insisting on an objective notion of truth and knowledge while acknowledging at the same time that the spirit of time (*Zeitgeist*) is more on the left. Just as in the case of physics where "the possibility of knowledge of the objective states of affairs is denied and it is asserted that we must be content to predict results of observation", in mathematics "many or most mathematicians denied that mathematics represents a system of truths" (Ibid. 377). Obviously Carnap will side with the left in Gödel's schema. This peculiar situation makes the comparison between them a certain paradigmatic interest for philosophy in general, above their own particular ideas and the particular topic of philosophy of mathematics.

1.2 What Is the Syntactical Interpretation of Mathematics [25] (SIM)?

1.2.1 The Mathematical Part

In a nutshell, SIM in the technical sense refers to the view which interprets mathematical propositions as true expressions solely by virtue of the syntactical conventions governing the use of those expressions, and not in terms of any objective mathematical fact. Their truth is therefore on a par with the true statement that "all mares are horses", simply because we choose the convention to use the term "mare" as an abbreviation for "female horse". These conventions are called "syntactical" (linguistic or formal) because they don't refer to any extra-linguistic objects but solely to the outward structure of the symbols. In Carnap's words,

> By the logical syntax of a language, we mean the formal theory of the linguistic forms of that language—the systematic statement of the formal rules which govern it together with the development of the consequences which follow from these rules.
>
> A theory, a rule, a definition, or the like is to be called formal when no reference is made in it either to the meaning of the symbols or to the sense of the expressions, but simply and solely to the kinds and order of the symbols from which the expressions are constructed. (Carnap 1937, 2)

[25] As Gödel made it clear, the terms "mathematical", "mathematics" in his discussion are used as synonymous with "logico-mathematical", "logic and mathematics", following Carnap, but without the implication that no borderline between the two can be drawn.

For example, the usual logical axiom "X→XVY" and the mathematical theorem "2+2=4" will be interpreted respectively as that "Every statement of the form XVY is a direct consequence of X and the expression '2+2' and '4' are always mutually substitutable". In §4 of the third version of "Is mathematics syntax of language", which is the richest in content of all the six versions, Gödel mentioned Carnap's effort in §34 of *LSL*[26], Ramsey's attempt to reduce mathematics to tautologies and Hilbert's program as those who really carried this syntactical program out.[27] The possibility of the syntactical program is largely due to the development of the axiomatic method in modern logic and the foundation of mathematics, i.e., nearly all of classical mathematics (although not all due to incompleteness) and logic can be incorporated in a formal system with just a few primitive axioms (or axiom schemata) and rules of inference. Besides, this formal system of mathematics seems to form a separate part from other empirical sciences in that no other empirical propositions follow from them. This feature, together with the psychological fact that propositions of logic and mathematics seem to not share the same status of truth as empirical ones in that no possible state of affairs are excluded by them, adds to the belief that a separate

[26] This book may well be seen as the *magnum opus* of Carnap . It is the culmination of Carnap's thoughts on logic, mathematics and philosophy before 1937 and the stepping stone for modification in later years. As one of the earliest books about this topic, Carnap presents in the book the state of the art in metamathematics in a systematical way, including discussions of the ω-rule, the arithmetization of syntax, the incompleteness theorems, the inexhaustibility of mathematics and the fix-point lemma, the logical and semantic paradoxes and the undefinability theorems, as well as set theory and Skolem's paradox, among many other things. For the historical background leading to this work, see (Awodey and Carus 2009; Uebel 2009).

[27] See (Ramsey 1925; Hilbert 1925). Ramsey's aim is to reduce all of mathematics to explicit tautologies a=a by definitions alone admitting propositions of infinite length, thus extending Wittgenstein's idea of tautology from logic to the whole of mathematics. Hilbert's formalistic foundation of mathematics can be seen as a special elaboration of SIM because axioms and rules of inference of a mathematical system, due to its formal character, can be interpreted by syntactical rules which stipulate that formulae of certain structures are true (axioms) and all formulae obtained from some other given formulae by certain formal operations are true, and nothing else is true.

treatment must and can be given for mathematics.

1.2.2 The Philosophical Part

Despite the similarity of the technical developments, the philosophical positions of these authors who developed the program (Carnap, Ramsey and Hilbert) are totally different, and it's in Carnap's work that the original purpose and the chief philosophical interest of SIM is most clearly articulated. The main philosophical problem for Carnap, qua logical empiricist who believes that all knowledge, in the last analysis, must be based on experiences (or perceptions), is to reconcile strict empiricism with the a priori certainty of mathematics. The alternatives to base mathematical knowledge either on pure intuition/reason like Kant or on empirical generalizations like Mill are all unsatisfactory for Carnap. It is Wittgenstein's view as expressed in the *Tractatus* (Wittgenstein 1922) that all logical propositions are analytic—in the specific sense that they hold in all possible cases and therefore do not have any factual content—that appealed to Carnap. By extending Wittgenstein's view to higher logic (including mathematics), Carnap recalled, "it became possible for the first time to combine the basic tenet of empiricism with a satisfactory explanation of the nature of logic and mathematics" (Carnap 1963, 47). Along with this view about the nature of mathematical truth and the precise demarcation of formal and factual sciences is the radical view that formal sciences (including mathematics and logic) do

not have any objects at all and they are merely expedient auxiliary instruments:

> These auxiliary statements, namely, the analytic ones… have indeed no factual content or, to speak in the material idiom, they do not express any matters of fact, actual or non-actual. Rather they are, as it were, merely calculational devices, but they are so constructed that they can be subjected to the same rules as are the genuine (synthetic) statements. In this way they are easily applicable devices for operations with synthetic statements. (Carnap 1935, 126)

The same position is emphasized again in *LSL* as the thesis that mathematics is void of content: "the latter, the so-called 'real' sentences, constitute the core of science; the mathematico-logical sentences are analytic, with no real content, and are merely formal auxiliaries." (Carnap 1937, xiv) And later, "an analytic sentence is not actually 'concerned with' anything, in the way that an empirical sentence is; for the analytic sentence is without content". (ibid. 7)

It's true that Carnap in his later philosophical works, especially (Carnap 1950a) would no longer hold on to such formulations about the differences between mathematico-logical sentences and empirical sentences[28]; however, as Gödel pointed

[28] Nevertheless, even in the framework of that work where Carnap famously proposed the distinction between internal and external questions, the opposition between formal and factual propositions where only the latter can be said to have content cannot be valid since existence of physical objects then becomes external questions which has no cognitive content at all. That is to say, even from the empirical standpoint, there is no reason to answer the

out clearly, he was not

> …concerned with a detailed evaluation of what Carnap has said about the subject, but rather my purpose is to discuss the relationship between syntax and mathematics from an angle which, I believe, has been neglected in the publications about the subject. For, while the syntactical program itself and its elaboration, as far as it is possible, have been presented in detail the negative results as to its feasibility in its most straightforward and philosophically most interesting sense have never been discussed sufficiently. (Gödel 1953/9, 336)

Apparently, rather than the historical question of a critical assessment of Carnap's philosophy, Gödel is more concerned with the conceptual question concerning the relationship between syntax and mathematics, where Carnap is taken to be the best representative of the syntactical view. The most interesting philosophical aspects of SIM from a realistic point of view, as Gödel formulated them, are as follows:[29]

> (1) Mathematical intuition, for all scientifically relevant purposes, in particular for drawing conclusions as to observable facts occurring in applied mathematics, can be replaced by conventions about the use of symbols and

question of the objective existence of mathematical and empirical objects differently.

[29] Gödel formulated them a little bit differently in version three and five, see (Gödel 1953a, 337) and (Gödel 1953b, 356). The following is a combination of the most important common points.

their application.

(2) Being merely consequences of conventions about the use of symbols, mathematics is known a priori without appeal to any intuition, compatible with all possible experiences, thus void of content and cannot be disproved by experience, in contradistinction to other empirical sciences which describe certain objects and facts.

These two assertions are not independent but closely related, because "if the prima facie content of mathematics were only a wrong appearance, it would have to be possible to build up mathematics satisfactorily without making use of this 'pseudo' content" (ibid. 346). Thus, (2) implies (1) but not vice versa, for it is conceivable that some part of mathematics, even if it can be reduced to syntax, still has an objective content. This, of course, doesn't exclude the possibility that we have independent reasons to show the falsehood of (2) apart from the evidence of the falsehood of (1). The basic aim for Gödel in his paper is to show that under the ordinary interpretations of the terms occurring in the above two assertions, which seems to be necessary if SIM is to serve its philosophical purpose, both can be shown to be false. He even goes as far as saying that SIM is "refutable, as far as any philosophical assertion can be refutable in the present state of philosophy" (Goldfarb 1995, 325). It is to the thorough examination of those arguments and their possible objections that we turn in the section below.

1.3 Gödel's Main Arguments Against SIM and Possible Objections

In this section we will first discuss the simplest version of SIM, due to its relation to the notion of "analyticity", and then discuss claim (2) above, that mathematics is void of content before turning to the more delicate question of claim (1) and consistency proofs.

1.3.1 A Simple Version of SIM and Analyticity

In the Gibbs lecture Gödel launched a brief yet decisive argument against the simplest version (also maybe the most common known and popular one) (Gödel 1951, 316). Since the most common type of linguistic conventions are definitions (either explicit or contextual), this version of SIM consists in the assertion that mathematical propositions are true solely owing to the definitions of the terms occurring in them. More precisely, according to this version of SIM, by successively replacing all terms by their definientia, any theorem can be reduced to an explicit tautology of identity like a=a. This notion of the analytic nature of mathematics is typical of most logical positivists in their early period, mainly under the influence of Wittgenstein. However, this view is demonstrably false even for number theory, since such a reduction would yield a decision procedure for the truth and falsehood of every mathematical proposition, which cannot exist due to the undecidability theorems by Church and

Turing[30] (Church 1936b; Turing 1936). An interesting point here about this simple version of SIM is its relation to the notion of "analyticity". Apart from this narrow purely formal sense of analyticity that reduces axioms and theorems of mathematics to special cases of law of identity and disprovable propositions the negation of this law via substituting definiens for the definiendum occurring in those propositions, Gödel suggests that there is a second sense in which "a proposition is called analytic if it holds 'owing to the meaning of the concepts occurring in it', where this meaning may perhaps be undefinable, (i.e., irreducible to anything more fundamental)" (Gödel 1944, 139). The two senses, according to Gödel, would be better called "tautological" and "analytic". We will see the consequence of this distinction in the discuss about the content of mathematics soon. Having rejected this simple version of SIM, a more refined one then has to aim at proving that every demonstrable mathematics proposition can be deduced from certain syntactical conventions. We will see below that even in this weakened form SIM fares no better.

1.3.2 Does Mathematics Have Content?

Gödel has two different lines of arguments against the view that mathematics is void of content while empirical sentences are not. The first is that Carnap's definition of content is artificial and arbitrary. Using Carnap's method in a different way we can make any category of sentences have no content at all or have full content. The second

[30] Thus, it is really surprising to find that as late as 1945 leading positivist as Hempel still adhered to such a wrong view. See (Hempel 1945).

argument shows that the view mathematics has no content is a result of asymmetrical treatment of different parts of a theory. It contains two aspects: only laws of nature together with mathematics (or logic) have verifiable consequences and that mathematics can add some content to a physical theory, i.e., mathematics is not always conservative over empirical sentences. This sounds very similar to Quinean holism and the indispensability argument for mathematical realism. However, closer scrutiny shows that there are striking differences between Gödel's argument and Quine's.

1.3.2.1 Arbitrariness of Carnap's Definition of Content

The notion of sameness of content, in connection with the notion of meaning, is notoriously difficult, which can be seen in Carnap's attempt to define analyticity through them and Quine's sustained criticism.[31] Intuitively we would think logical equivalence as a good criterion of the sameness of content, as can be seen from the mathematical example: one would consider the statement "There doesn't exist an even natural number greater than 2 which cannot be decomposed into the sum of two prime numbers" and the statement "All even natural numbers greater than 2 can be decomposed into the sum of two prime numbers" to be formulations of the same mathematical claim. And yet suppose that we have an axiom system with an initial finite of set of axioms Σ and two totally different theorems P and Q that are provable from the axioms. Then we would hardly be ready to say that the claim "P follows

[31] See (Quine 1951a) and (Carnap 1952) for example.

logically from Σ" has the same sense as "Q follows logically from Σ" even when both are logically true and equivalent. More generally, we will declare two assertions of theoretical physics to have the same sense when one is obtained from the other by a transformation of a mathematical expression it contains, but this is not permissible in general with mathematical assertions. We will say of a formulation of a mathematical proposition that its sense is not changed by an elementary logical transformation, but this will no longer hold when the elementary logical relations themselves are considered. These general considerations suggest to us that the notion of content is very likely to be relative or comes in different levels and what is regarded as the content of a proposition largely is a question of what one is interested in. For example, consider the statement "It will be sunny or it will not be sunny tomorrow", one may very well say against the positivists, that although it says nothing about the weather tomorrow, it does express a property of "not" and "or". Gödel's first main argument against the view that mathematics has no content is that "the reasoning which leads to the conclusion that no mathematical facts exist is nothing but a petitio principii, i.e., 'fact' from the beginning is identified with 'empirical fact', i.e., 'synthetic fact concerning sensations'" (Gödel 1953a, 351). In order to bring out this argument more clearly, we need to see how Carnap defends this thesis in *LSL*.

For Carnap, the fundamental difference between mathematical truths and empirical truths, as we mentioned earlier, is that the latter are synthetic and have content while the former are analytic, with no real content, and purely formal. In order to support this claim Carnap elaborates a delicate definition of analyticity, making it on the one hand

a complete[32] formal criterion of validity for mathematics and logic and on the other hand the distinguishing feature between logico-mathematical truth and empirical ones. It is thus to this essential concept that we turn to in order to see whether Carnap succeeds in achieving his aim.

There are actually two approaches used by Carnap to define the notion of analyticity and other related c-concepts.[33] The first approach occurs in his discussion of specific languages—language Ⅰ and Ⅱ, where he starts with a division of primitive terms into 'logical'[34] and 'descriptive', upon which he defines the notion of analytic, synthetic and other c-notions. The second approach occurs in the discussion of general syntax, where he starts with an arbitrary relation of direct consequence and uses it to define other c-notions and then give a classification of terms into 'logical' and 'descriptive'.[35] In both cases, the content of a sentence is defined to be the set of its non-analytic consequences. It then follows immediately from the definition that logico-mathematical sentences are analytic or contradictory and without content (assuming consistency).

But the problem is that in both the first and the second approach the distinction between logical and extra-logical terms or which consequence counts as a proper

[32] That is to say, the definition of analyticity need to have the consequence that every mathematical proposition is either analytic or not-analytic, i.e., contradictory. Using our modern terminology, what Carnap needs is actually some sort of truth definitions for mathematical propositions.

[33] Consequence concepts, most of which we would deem as semantic today such as logical consequence and equivalence which could well be non-recursive (indefinite, in Carnap's words), in contrast to d-concepts, derivation concepts such as derivable, demonstrable, refutable which are usually recursive or semi-recursive.

[34] Carnap's use of 'logical' is wider than what we use today, i.e., not only logical but also mathematical such as symbols from arithmetic and set theory.

[35] To be more precise, a statement is analytic if it is a consequence of the null class of statements, contradictory if every statement of the language is a consequence of it, and determinate if it is either analytic or contradictory. A statement is synthetic if it is indeterminate, i.e., neither analytic nor contradictory. The main idea for a classification of the term is that "the formally expressible distinguishing peculiarity of logical symbols and expressions [consists] in the fact that each sentence constructed solely from them is determinate" (Carnap 1937, 177).

logical consequence (provable by first order logic or in a first order arithmetical theory) is totally arbitrary and no principled distinction and independent justification are made. For example, if one calls only the primitive terms of first-order logic "logical" and calls the extended terms of arithmetic and set theory "extra-logical" then we will come to the conclusion that only theorems of first-order logic are without content while statements of arithmetic and set theory have content. In the case of consequence it doesn't fare any better. Let S be first-order Peano Arithmetic and for the direct consequence relation take "provable in PA", then any undecidable sentence such as the statement stating the consistency of PA will be deemed descriptive, synthetic and having content. On the other hand, by using the relation of "consequence due to laws of nature", a concept of "accidental content" could be defined, according to which laws of nature would be "void of content" and only those facts not subject to any natural laws have content. What Carnap has done really is just to provide an *ad hoc* piece of technical machinery to yield the results he desired, without, however, giving any independent argument or justification for the original choice. This shows that Carnap's definition, at best, is only a sleight of hand; as Peter Koellner put it, "it is trivial to prove that mathematics is trivial if one trivializes the claim" (Koellner 2009b, 89). The situation is exactly the same in Tarski's first proposed definition of logical consequence, which is based on a priori classification of all terms into logic and extralogical. Tarski concluded his essay with his customary caution towards philosophical questions:

> Underlying our whole construction is the division of all terms of the language discussed into logical and extra-logical. This division is certainly not arbitrary. … On the other hand, no objective grounds are known to me which permit us to draw a sharp boundary between the two groups of terms. It seems to be possible to include among logical terms some which are usually regarded by logicians as extra-logical without running into consequences which stand in sharp contrast to ordinary usage. In the extreme case we could regard all terms of the language as logical. The concept of formal consequence would then coincide with that of material consequence. The sentence X would in this case follow from the class K of sentences if either X were true or at least one sentence of the class K were false. (Tarski 1936, 213–14)

Tarski quite clearly realized the importance of the question of demarcation, not only for logic, but also for epistemology and concepts like “analytic” and “tautological”, but he remained a skeptic towards these problem:

> I consider it to be quite possible that investigations will bring no positive results in this direction, so that we shall be compelled to regard such concepts as ‘logical consequence’, ‘analytical statement’, and ‘tautology’ as relative concepts which must, on each occasion, be related to a definite, although in greater or less degree arbitrary, division of terms into logical and

extra-logical.[36] (ibid.)

Gödel would disagree with Tarski (and after him, Quine) on this matter. He is of the opinion that the distinction between analytic and synthetic is absolute, which is in a way more Carnapian. However, in order to refute Carnap's thesis that mathematics has no content, the arbitrariness argument would suffice.

1.3.2.2 Asymmetry Argument

The first aspect of the argument refers to the role mathematics plays in the construction of a theory. Using Gödel's own words, "laws of nature without mathematics are exactly as 'void of content' as mathematics without the laws of nature. The fact is that only laws of nature together with mathematics (or logic) have consequences verifiable by sense experiences. It is, therefore, arbitrary to place all content in the laws of nature."[37] (Gödel 1953a, 348–49) That is to say, while only physical theory and mathematics together have observational consequence, the positivists, by treating them in an asymmetrical way, tend to ignore the formal part and attribute the content exclusively to physical laws. The striking resemblance of the wording of these expressions with Quinean holism and the "Quine-Putnam

[36] This attitude is again reflected in one of Tarski's letters in 1944, which seemed to anticipate Quine's later famous attack on the distinction of analytic/synthetic. See (White and Alfred 1987).
[37] Gödel did mention the possibility that some laws of nature might have verifiable consequences due to the rules for the use of symbols in them, but (1) this is not true for all laws of nature and (2), in general not all verifiable consequences of a law of nature can be obtained in such a way. (Gödel 1953a, 349 footnote 37).

indispensability argument"[38] makes it very easy to slip into an extreme holism which blurs and eventually denies any essential distinction between the logico-mathematical part and the physical part of a scientific theory: they are not different in kind, but only in degree, reflecting our pragmatic preferences in revising a theory against the accumulation of new observational evidence. However, this cannot be further from what Gödel would say about the structure of theory. To start with, Gödel's argument was a positive one to show that mathematics does have content, at least if we allow physics to have one, whereas in Quine's case, although not strictly incompatible with either his holistic empiricism or his naturalized epistemology, to admit the existence of abstract objects in general and mathematical objects and facts in particular is ultimately a grudging choice, one that he very much wanted do without: there is ultimate bias and asymmetry in Quine's argument just like Carnap's. As Burgess puts it,[39]

> [It] does really seem to be Quine's view that only the *indispensable necessity* of positing mathematical entities in science can justify belief in them; the mere fact that such posits are a *customary convenience* is not enough for Quine, as it might be for another philosopher. There is a bias against mathematical entities in Quine's thinking, and he never doubts that it

[38] See (Putnam 1971). However, there is no canonical source for the so-called argument in Quine's writings, comparable to Putnam's little book, but merely scattered passages here and there in Quine's papers from the 1950s or 1960s, see (Quine 1951a, 1951b, 1960, 1971). The current discussion of the indispensability argument, in its more precise form, usually differs from Quine's text in one way or another.

[39] Burgess's argument goes further in suggesting that not only the bias towards abstract objects including mathematical ones, but also the opposition of "ontology" against "ideology" in Quine's philosophy does not strictly follow from the central tenets of Quine. See (Burgess 2013, 293).

would be better to get rid of them if only we could. Nominalism would be his position if he could make a go of it. (Burgess 2013, 291)

Moreover, the greater difference can be seen from Gödel's double attitude towards the SIM: although SIM is wrong in thinking that mathematics has no content, this view "doubtless has the merit of having pointed out the fundamental difference between mathematical and empirical truth. This difference … is placed in the fact that mathematical propositions, as opposed to empirical ones, are true in virtue of the *concepts* occurring in them" (Gödel 1953b, 357). What is wrong with SIM is that it interprets the meaning of the mathematical concepts as something man-made and originating from our conventions, while the truth, for Gödel, is that "these concepts form an objective reality of their own, which we cannot create or change, but only perceive and describe" (Gödel 1951, 320). This is one of the most striking passages Gödel had ever written about his explicit conceptual realism. Despite the fact that he never proposed a mature theory about concepts[40], Gödel did provide us with some examples that can hardly be doubted, such as the principle of mathematical induction for the natural numbers, the axiom of separation of set theory applied to the theory of integers and the inference *modus ponens*. If we are not going to be arbitrary, then the truth of modus ponens, for example, is just like the truth of a perception such as "This is red", the difference between which lies only in the fact that the former case is a

[40] Maybe due to the unsolved intensional paradoxes connected to concepts as individual existent entities, see the chapter on Gödel and Russell for more details about this point.

relation between concepts alone while the latter is a relation between objects and concepts. It is exactly in regards to this similarity between perceiving truth about concepts and truth in ordinary perception, both "[forcing] themselves upon us as being true" (Gödel 1964, 268) that Gödel speaks of the existence of mathematical intuition that resembles physical sense, which enables us to perceive certain intuitive mathematical and logical truths based on the meaning of concepts occurring in them; it also allows us to form an idea of abstract objects based on the impressions given in mathematical intuition. Although the data of mathematical intuition cannot be associated with actions of physical things upon our sense organs, it need not be something subjective as Kant asserted, "rather, they, too, may represent an aspect of objective reality, but as opposed to the sensations, their presence in us may be due to another kind of relationship between ourselves and reality" (ibid.). The existence of an objective content of mathematics can be regarded as a confirmation of this view about the nature of mathematics. What's more, the assumption of mathematical intuition and mathematical truth as conceptual truth can satisfactorily solve the problem of the apparent intuitiveness of certain parts of mathematics (such as elementary arithmetic) as opposed to the rest, a problem Quine and the indispensability argument can hardly be said to deal with adequately.[41] To view mathematical truth as truth about mathematical concepts independent and separate from empirical truth also makes it possible to see what particular significance mathematics can contribute to physics in a scientific theory, rather than suggesting a vague metaphor of web of beliefs. Gödel is

[41] Charles Parson, among others, has emphasized this point to a great extent, see for example (Parsons 1980).

quite articulate about the different roles that mathematics and physics play in a theory:

> What mathematics adds to the physical laws, it is true, are not any new properties of physical reality, but rather properties of the concepts referring to physical reality—to be more exact, of the concepts referring to combinations of things. But such properties are something quite as objective as properties of physical reality and even verifiable by sense experience under the hypothesis that certain laws of nature which can be confirmed independently of mathematics hold good. (Gödel 1953a, 349)

In order to understand better the relationship between mathematics and laws of nature, we need to come to the second aspect of the asymmetry argument, namely mathematics is not conservative over empirical sentences, but can add some new content to the laws of nature, thus refuting the view that only laws of nature have content while mathematics doesn't. The first difficulty in making this idea clear is that it is difficult to even make precise the claim that mathematical systems are conservative with respect to the empirical language since much of that language is already laden with mathematics.[42] Thus, it doesn't make absolutely clear sense when we say mathematics is conservative over laws of nature, hence void of content. Even disregarding this difficulty, we can still refute this claim by considering the following

[42] As Howard Stein argues, the real problem for Carnap's philosophy is that he sticks to the idea of an "observational" part of any language: "we have no language at all in which there are well-defined logical relations between a theoretical part that incorporates fundamental physics and any observational part at all—no framework for physics that includes observational terms, whether theory-laden or not. (Stein 1992, 290). Bernays mentions a similar point in his paper about Carnap's philosophy, especially as represented in *LSL,* see (Bernays 1961).

example.[43]

> In classical mechanics, it is possible to set up a system of perfect elastic reflectors and set a point mass bouncing among them in such a way that one can encode the halting problem. More precisely, for any Turing machine M, one can encode the question "Will M halt on null input?" by the question "Will the particle pass through point p of the output screen?" To be sure, this form of the question involves a minimal amount of mathematics, but it seems clear that it is an idealized empirical question. Let M be a Turing machine that halts on null input if and only if ZF+AD is inconsistent. Now take the arrangement of perfect reflectors to encode this version of the halting problem, and set the question as "Does the particle ever pass through the halting point p of the output screen?" Then Mathematical systems are not conservative with respect to this empirical question.

In the above scenario, it's obvious that the answer to an empirical problem depends on the answer of abstract set theory.[44] It is also conceivable in cases that the solution of certain mathematical problems might also solve problems formerly undecidable in mathematical physics, thus leading to new empirically verifiable consequences just like new laws of nature.

---

[43] This example comes from (Koellner 2009a), which he again attributes to (C. Moore 1990, 1991). For the philosophical implications of such examples, see (Pitowsky 1996).

[44] The answer is undecidable in PA, but settled negative in much stronger system of set theory. Most set theorists would consider ZF plus the axiom of determinacy to be consistent, thus the answer to be no.

So both arguments undermine Carnap's claim that mathematics, as opposed to empirical sentences, is void of content. The neglect of conceptual content as against factual content is not justifiable from either considerations of its own (ignoring the similarity between mathematical intuition and perception), or its consequences in scientific theories (mathematics is not always conservative over laws of nature). That mathematics does have content also results from the failure of syntactical interpretation itself, on which we will focus below.

1.3.3 *Petitio Principii* and the Epistemological Significance of Consistency Proofs

In this section, I will first discuss Gödel's argument (3.3.1), which is a type of general *petitio principii* (circular argument), and then discuss two possible objections in defence of Carnap. The first objection (3.3.2) comes from Awodey and Carus, concerning the necessity of demonstrable consistency rather than dependable consistency. They argue that Gödel's requirement for SIM to be demonstrably consistent is unreasonable. I will then first discuss (3.3.2.1) the general philosophical importance of consistency proofs generated by Gödel's second incompleteness theorem, its relevance to skepticism toward general mathematics and its applicability to Carnap's case (3.3.2.2). The second objection (3.3) comes from more extrinsic rather than intrinsic considerations regarding the whole issue. Goldfarb and Ricketts urge us to view SIM not as an epistemological thesis but a pragmatic proposal, traditional foundational problems having transformed themselves in the hands of

Carnap. Under this new point of view, many of the charges by Gödel against Carnap will disappear, but on the other hand Carnap cannot persuade Gödel either: there is no real confrontation here, but only a stand-off. I will first point out the inconclusiveness of this argument and show the inadequacy of this interpretation by several different arguments.

1.3.3.1 Petitio Principii and the Non-feasibility of the Syntactical Program

Gödel's core argument against SIM comes from his own incompleteness theorem concerning the impossibility of a consistency proof within the formal system itself. In order for SIM to achieve its proclaimed philosophical aim and in order to fulfill the conditions for incompleteness theorem to apply, Gödel stated six conditions restricting SIM. They are the following (Gödel 1953/9a, 337–41):

(1) Since SIM aims at dispensing with mathematical intuition without impairing the usefulness of mathematics for empirical science, it will have to be required that mathematics is covered to the full extent, namely, it must refer to all classical mathematics.

(2) "Language" will have to mean some symbolism which can be actually exhibited and used in empirical world. In particular, sentences have to be finitely long.

(3) Rules of syntax have to be finitary, i.e., they must not contain phrases referring to an infinite totality.

(4) Rules of syntax must not imply the truth or falsehood of any "factual" sentence (i.e., one whose truth, owing to the semantical rules of the language, depends on extralinguistic facts[45]), thus they must be demonstrably consistent.

(5) Not only must formal axioms and proof procedures be deduced from suitably chosen rules of syntax, but also the conclusions as to ascertainable facts which are obtained by applying mathematical theorems and which formerly were based on the intuitive truth of the mathematical axioms must be justified by syntactical considerations. This justification, however, requires also a proof of the consistency of the syntactical rules.

(6) Not only rules of syntax, but also in the derivation of mathematical axioms from them and in the proof of their consistency, only finitary syntactical concepts (i.e., concepts referring to finite combinations of symbols) and procedures of proof based on these concepts can be applied.[46]

[45] In his introduction to Gödel's paper (Goldfarb 1995), Warren Goldfarb raised an objection that Gödel's argument presupposed the existence of a realm of 'factual' or the 'empirical' being available in advance, independently of and prior to the envisaged rules of syntax. This presupposition is foreign to Carnap and thus Carnap is immune to Gödel's objection. The weakness of Goldfarb's objection can be seen easily from the additional comment here, where Gödel puts factual in quotation marks and stressed that "factual" in terms only of semantical rules.

[46] In a footnote Gödel mentioned the relationship between syntax satisfying all the six conditions and Hilbert's finitism: "I believe that what must be understood by 'syntax', if the syntactical program is to serve its purpose, is exactly equivalent to Hilbert's 'finitism', i.e., it consists of those concepts and reasoning, referring to finite combinations of symbols, which are contained within the limits of 'that which is directly given in sensual intuition'", and he believed that the limit of syntax thus defined "is equivalent (by a one-to-to correspondence of its objects) with recursive number theory" (Gödel 1953a, 341 footnote 19). For more about the discussion of Gödel's hesitation about the exact limit of finitism, see the chapter on Gödel and Hilbert.

Although versions of SIM which conform to several but not all of these conditions have been elaborated and are actually possible[47], all six conditions have to be satisfied, however, if SIM is to serve the epistemological purpose for reducing mathematics to symbolic conventions. The central problem, the consistency of those syntactical rules in compliance with finitary means, however, poses a stumbling block for every such attempt. All the actual elaborations of SIM have failed because Ramsey's ideas necessitate admitting propositions of infinite length and Carnap uses non-finitary syntactical rules and arguments in his reduction and Hilbert's search for a finitary consistency proof for classical mathematics is also impossible due to Gödel's second incompleteness theorem. In all these attempts, abstract concepts [48] based on mathematical intuition or certain axioms about them, rather than just considerations about finite combinations of symbols have to be invoked. This suggests to us that the failure is not just a feature of these particular attempts, but that it has deeper reasons. Gödel formulated the general conclusion as "the non-feasibility of the syntactical program", or what amounts to the same, "non-eliminability of the mathematical content of an axiomatic system by the syntactical interpretation":

> The scheme of the syntactical program to replace mathematical intuition by rules for the use of symbols fails because this replacing destroys any

[47] For example, formalism under the requirements 2-6 has yielded a syntactical foundation only for a small part of mathematics, i.e., certain parts of mathematics (such as first order Peano arithmetic with the restriction of the induction schema applied only to quantifier-free formulas) can be reduced to syntactical conventions in terms of a finitary consistency proof.

[48] Such as the concept of "proof" and "function", understood not in terms of a sequence of expressions complying with certain formal restrictions or an expression of the formalism satisfying certain formal conditions, but "a sequence of thoughts convincing a sound mind" and "an understandable and precise rule associating mathematical objects with mathematical objects" (Gödel 1953a, 341 footnote 20).

> reason for expecting consistency, which is vital for pure and applied mathematics, and because for the consistency proof one either needs a mathematical intuition of the same power as for discerning the truth of the mathematical axioms or a knowledge of empirical facts involving an equivalent mathematical content. (Gödel 1953a, 346)

As to the particular case of Carnap, he violates the requirement (5), which means he uses non-finitary syntactical rules in his *LSL*. However, in Gödel's view, the finitary requirement "should be beyond dispute". For, since the aim of SIM is to replace mathematical intuition and assumptions of mathematical objects and facts, our considerations about the abstract and transfinite concepts of mathematics have to be based on considerations about finite combinations of symbols only. If for the formulation of syntactical rules themselves some of the very same abstract or transfinite concepts are being used—or in the consistency proof, some of the axioms usually assumed about them—then "the whole program completely changes its meaning and is turned into its downright opposite: instead of clarifying the meaning of the non-finitary mathematical terms by explaining them in terms of syntactical rules, non-finitary terms are used in order to formulate the syntactical rules" (ibid. 342), and the same for axioms in the justification of the syntactical rules as consistent.[49] That is to say, the idea to reduce all mathematics to conventional syntactical rules about the

[49] For example, in §34h, Carnap gave a proof of the axiom of choice (what he called "the principle of selection") but he did so only by assuming the principle itself in the syntax-language (or meta-language). Carnap, however, did not view this as anything circular as he insisted that the question as to whether axiom of choice is valid or not is not a theoretical problem to be decided, but only a matter of choice in terms of expedience. We will discuss this idea below.

use of symbols is actually just a false appearance in that mathematics must in some way or other be presupposed in order to set up the those rules, which again makes the original epistemological aim also as much of an illusion.[50]

1.3.3.2 Objection : Demonstrable Consistency or Dependable Consistency?

The first major objection against Gödel's argument is one from Awodey and Carus.[51] They don't doubt that a finitary consistency proof for syntactical conventions is impossible, rather they argue that it's not necessary, as Gödel insists. The defect in Gödel's argument, as they see it, is in condition (4) as listed above, where Gödel says that "this requirement not only follows from the concept of a convention about the use of symbols, but also from the fact that it is the lack of content of mathematics upon which its a priori admissibility in spite of strict empiricism is to be based. This requirement implies that the rules of syntax must be demonstrably consistent, since from an inconsistency every proposition follows, all factual propositions included" (ibid. 339). To see the fallacy, they rearrange Gödel's argument in a step-by-step

[50] This type of argument against the "linguistic view of mathematical and logical truth" can also be found in Quine, see (Quine 1936, 1960), which bears again a resemblance to Lewis Carroll's famous argument (Carroll 1895), in that all of them deny that a *purely* conventional account can be given for mathematical and logical truth, i.e., some truth or some inference rules must be taken as given before any such conventions can be made. However, there exists a difference here between Gödel's and Quine's arguments which makes Gödel's more penetrating than Quine's. Quine's argument against the conventional view of logical truth is that, conventions being finite in number while logical truth being infinite, we need logic again in deducing the infinite number of logical truths from the finite number of conventions, thus a circularity. In Gödel's argument, even disregarding the problem of the infinite number of logical truths, i.e., even if they are finite in number, a circularity still exists. Take the proposition "2+2=4", for example. We can indeed reduce this proposition into explicit tautologies by definition alone. However, in order for the "+" operation in this statement be in accord with the ordinary (and the intended) one, which can only be defined in the familiar contextual way, one has to use the concept of "finite manifold", which in fact is equivalent to arithmetic, thus again the circularity. See (Gödel 1951, 318–19) for an elaboration of this idea.

[51] The objection was raised in different papers with more or less the same argument, see (Awodey and Carus 2003, 2004, 2010).

way.[52] First we have the premises:

(A) SIM implies that mathematics is empirically vacuous.

(B) If a stipulated language for mathematics is inconsistent, then it may have empirical consequences.

From B by contraposition, it follows that

(B′) If a stipulated language for mathematics is to be empirically vacuous, it must be consistent.

From A and B′ it only follows that

(C) SIM requires the consistency of any stipulated language for mathematics.

And it doesn't follow, as Gödel would suggest that

(D) SIM requires the provable consistency of any stipulated language for mathematics.

[52] The following is my reformulation of the argument presented in (Awodey and Carus 2010, 268-269).

Only a variant of D is required by Gödel's argument, namely,

(E) A *proof* of SIM requires the *provable* consistency of any stipulated language for mathematics.

But this conclusion, that we cannot prove that mathematics is syntax of language by mathematics itself, rather than undermining SIM, is in complete harmony with it. For SIM implies the vacuity of mathematics, which would surely be violated if the viewpoint itself be proved mathematically, as such a result itself would be a non-trivial mathematical proposition.[53] And they also support their conclusion by an analogy with physical assertions: we cannot prove them in a strict way, but that doesn't compel us out of the realm of knowledge into mere opinions. What is required for Carnap is not demonstrable consistency, but only dependable consistency, or in other words, empirical consistency—consistency based on the fact that no contradiction has arisen so far. To be sure, Gödel doesn't treat SIM as a thesis that must be proved mathematically, other theses in the foundations of mathematics like Platonism or realism cannot be proved in this sense either. It is the particular nature of SIM which makes a consistency proof necessary. Before we give a satisfactory reply to this objection, we need to have some general remarks about consistency proof, its relevance and necessity to many philosophical problems.

[53] This conclusion, I think, doesn't strictly follow from Awodey and Carus' arguments, for mathematics to be contentful is compatible with its empirical vacuity, i.e., nothing empirical follows from mathematics alone. In order for their argument to work, "empirically vacuous" in premise (A) has to be replaced by "vacuous tout court".

1.3.3.2.1 The Lightness and Heaviness of Consistency Proofs

Paul Rosenbloom attributes to Andre Weil the saying that "God exists, since mathematics is consistent, and the Devil exists, since we cannot prove it." (Rosenbloom 1950, 72) Gödel, however, is the missing link, for he supposedly proved, roughly speaking, that if mathematics is consistent then we cannot prove it. Similar remarks with a skeptical tune can be found in Nagel and Newman's popular book on Gödel's theorem. They maintain that Gödel for having proved that

> It is impossible to establish the internal logical consistency of a very large class of deductive systems—elementary arithmetic, for example—unless one adopts principles of reasoning so complex that their internal consistency is as open to doubt as that of the systems themselves. (Nagel and Newman 1958, 6)

The ease with which we invoke Gödel's theorem concerning consistency proofs to support some form of skepticism concerning mathematics can cause us to ignore an important distinction. This is the distinction between the degree of skepticism or confidence regarding mathematical axioms or methods of reasoning is justifiable or reasonable, and the bearing Gödel's theorem has on the matter. Under closer inspection it actually turns out to be that these two elements are not so related as they are usually

supposed to be. We might even come to the conclusion that Gödel's theorem is irrelevant to such skepticism. As Torkel Franzén puts it:

> So if we have no doubts about the consistency (or even stronger, the knowledge of the truth of the axioms) of ZFC, there is nothing in the second incompleteness theorem to give rise to any such doubts. And if we do have doubts about the consistency of ZFC, we have no reason to believe that a consistency proof for ZFC formalizable in ZFC would do anything to remove those doubts. (Franzén 2005, 105–6)

The often undue weight we tend to put on consistency proofs certainly comes from the strong influence of the philosophical legacy of Hilbert. It is Hilbert who first realized the importance of consistency proof in the debate about the foundations of mathematics; he turned this problem from a somewhat epistemological question into a precisely defined mathematical question, along with his basic philosophical considerations.[54] Hilbert tried to prove the consistency of strong theories like an axiomatic system for classical analysis or ZFC on the basis of very weak mathematical assumptions and finitistic reasoning, without assuming the existence of infinite objects and making only restricted use of logical principles, which are considered to be more secure and evident than those problematic ones used in the strong theories. In view of this observation, skepticism toward the formal systems we use in mathematics is only

[54] For more about the importance of consistency proofs in Hilbert's philosophy and the thorny problem of the extent of "finitism", see the chapter on Gödel and Hilbert.

warranted via Gödel's second incompleteness theorem together with the specific Hilbertian philosophical point of view that only finitistic reasoning embodies safe logical principles which is not open to doubt as regards its consistency. In other words, those who don't believe in the exclusively epistemological privileging of finitistic reasoning, just like Gödel who believes that some abstract reasoning can also be as intuitive as the concrete ones in Hilbert's finitism, can accept with equal confidence that there is no finitistic consistency proof for PA and observe that the consistency of PA is easily provable by other evident means. On the other hand, most consistency proofs in the logical literature are purely mathematical proofs independent of epistemological considerations, and always contribute a great deal more than just establishing mere consistency[55]. For example, Gentzen's proof (Gentzen 1936) for elementary number theory brings in the new idea of "ordinal analysis" and Gödel's proof (Gödel 1940) of the "relative consistency" of the axioms of choice in set theory, due to its constructive nature, establishes that every arithmetical theorem that can be proved using the axiom of choice can be proved without using that axiom—a fact that is by no means obvious, even given the consistency of ZFC. Even in the formalistic tradition itself this aspect of consistency proof was recognized gradually as the more important one. As Bernays wrote in 1961,

> The acknowledgment of the methodological importance of proof-theoretic investigations and in particular those concerning formal

[55] See (Kreisel 1958) for further examples and discussions.

> consistency, is not tied to the view that conventional, classical mathematics is dubious, or to the standpoint of "formalism" according to which classical mathematics is justified only as a purely formal technique. … The task of constructive consistency proofs is motivated by the high theoretical level that is present in classical mathematics. (Bernays 1961, 19–20)

All these considerations bring to the fore the fact that in most of the cases no particular significance or doubts should be attached to consistency proofs. Consistency proofs, being themselves a type of mathematical proof, become epistemologically significant only along with some prior philosophical assumptions. As Franzén nicely summarizes the situation:

> It is indeed perfectly possible to have doubts about the consistency of a theory T and to seek to eliminate those doubts through a consistency proof. In such a case we need to carry out the poof in a theory whose consistency is not equally open to doubt. However, a consistency proof may just as well be a perfectly ordinary mathematical proof … not aiming at allaying doubts about the consistency of mathematics, any more than proofs of arithmetical theorems in general are aimed at allaying such doubts. Gödel's theorem tells us nothing about what is or is not doubtful in mathematics. To speak of the consistency of arithmetic as something that cannot be proved makes sense only given a skeptical attitude towards ordinal mathematics in general.

(Franzén 2005, 112)

1.3.3.2.2 Consistency Proof in Carnap's Case

We have already seen from the discussion above that consistency proof is by no means a necessary condition for our confidence in mathematical theories. This conclusion might lead us to the same argument that Awodey and Carus bring about against Gödel. However in the particular case of Carnap, if SIM is to serve the philosophical purpose of showing that mathematics is void of content, then a consistency proof of the conventions is necessary.

> A consistency proof is indispensable because it belongs to the concept of a convention that one knows it does not imply any propositions which can be falsified by observation (which, in the case of the mathematical "conventions", is equivalent with consistency). Without a consistency proof the "convention" itself, since open to disproof, really is an assumption (or else laws of nature could also be interpreted to be conventions. (Gödel 1953a, 348–49)

Assumptions, like physical laws in theoretical physics do have content because they, unlike mere conventions where only questions of expedience can occur, can be

falsified. [56] If syntactical conventions are treated as assumptions capable of falsification, then they must have content too just like laws of nature. On the other hand, if only empirical inductive consistency is required, as Awodey and Carus have argued, then consistency must be interpreted to refer to the handling of physical symbols, which is empirically verifiable like a law of nature. However, if this notion of consistency is intended, then "mathematical axioms and sentences completely lose their 'conventional' character, their 'voidness of content' and their 'apriority' and rather become expressions of empirical facts." (ibid. 342) This simple but often neglected insight can be seen from examples in basic arithmetic. For example, if the operation of addition or multiplication is empirically known to be associative, then based on this empirical fact some conventions about dropping the unnecessary brackets may be adopted. But vice versa, from this convention the associativity of the operation, i.e., an empirical fact can be derived. In contrast to the view that mathematical conventions, apart from the knowledge about symbols necessary for setting them up, do not express or imply any facts, the above simple example reveals a general truth that conventions of mathematics are void of content only insofar as they add nothing to the mathematics which, before they can be made, must be known already. There are, of course, purely notational conventions which we can adopt either one way or the other; for example the choice between decimal and binary system for representing numbers, or the choice between sets and well-founded trees, determined

[56] Even from of the point of view of pure mathematics, there is a sense that mathematical propositions can be falsified, not by a false observational sentence, but by an inconsistency. It has to be admitted that consistent theories do have at least some "immanent existence", in contradistinction from an inconsistent one, just like a well-established physical theory against a wrong (falsified) one.

by the practical consideration of expedience relative to given purposes. But Carnap has gone too far in claiming all mathematical conventions to be of this kind. On the contrary, most of the interesting and important conventions in ordinary mathematics including his own syntactical rules are of the other kind, namely, the admissibility of which presupposed a certain system of knowledge. In a sense we cannot speak of conventions and their voidness of content in an absolute sense, since those conventions will at least presuppose the indispensable knowledge for making any linguistic conventions at all, which amounts to finitary combinatorics. Carnap's use of syntactical conventions extend to all classical mathematics and are in no way restricted to this restricted body of knowledge. Anyway, either the syntactical conventions in SIM need a consistency proof or they will lose their purely "conventional" character and become expressions of empirical facts, both of which are a direct refutation of the objections from Awodey and Carus.

1.3.3.3 SIM: Foundation or Proposal?

1.3.3.3.1 Pluralism and Tolerance

The second major objection comes from Goldfarb and Ricketts (Goldfarb and Ricketts 1992). They argue that Gödel had not understood the novelty of Carnap's *LSL*. Gödel had continued to consider Carnap's work in a foundationalist and reductionist perspective, but such foundationalism was abandoned by Carnap in *LSL*:

> It is a mistake to take Carnap be to trying to give an informative answer to the Kantian question 'How is mathematics possible?' or indeed to any similar question. What we are suggesting is that Carnap does not take the general clarification of the status of mathematics which *LSL* provides as being at all foundational, as addressing the issues that concerned not just Kant, but also his immediate predecessors Frege, Russell, and Hilbert … Rather, for the Carnap of *LSL*, questions of foundations, upon clarification, wind up being questions of what can be done inside various linguistic frameworks, or what sort of frameworks can be made usable. (Goldfarb and Ricketts 1992, 68)

By turning philosophical problems into the logical syntax of language[57], and, in particular by treating philosophical controversies as questions about the choice between certain particular language forms, for example, whether to include the law of excluded middle in the language or not, the debate in the foundation of mathematics becomes a pragmatic rather than a theoretical problem. In such a way Carnap can finally put an end to the "pseudo-problems and wearisome controversies" (Carnap 1937, xiv–xv). Related to such a transformation of the nature of philosophical problems is the famous pluralistic conception and principle of tolerance in *LSL*:

[57] See part V of *LSL* for a detailed account. For a critique of this "minimalist conception of philosophy", see (Koellner 2009a, 2009b).

> Our attitude to requirements of this kind is given a general formulation in the *Principle of Tolerance: It is not our business to set up prohibitions, but to arrive at conventions*. … *In logic, there are no morals.* Everyone is at liberty to build up his own logic, i.e., his own form of language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments. (Carnap 1937, 51–52)

In his autobiography Carnap claimed that the chief interest and motivation of *LSL* was to eliminate the search for the unique "correctness" in those fruitless philosophical debates:

> I wished to show that everyone is free to choose the rules of his language and thereby his logic in any way he wishes. This I called the "principle of tolerance"; it might perhaps be called more exactly the "principle of the conventionality of language forms". As a consequence, the discussion of controversies need only concern first, the syntactical properties of the various forms of language, and second, practical reasons for preferring one or the other form for given purposes. In this way, assertions that a particular language is the correct language or represents the correct logic such as often occurred in earlier discussions, are eliminated, and traditional ontological problems, in contradistinction to the logical or syntactical ones, for example,

problems concerning "the essence of number", are entirely abolished. (Carnap 1963, 54–55)

There is indeed lots of textual evidence which supports this non-substantive[58] interpretation of Carnap's position, in which Carnap is not really addressing any epistemological assertions about the foundations of mathematics and is only offering a proposal to view mathematics and logic in a new way. For example he has little concern for strictly finitary or constructive consistency proofs in *LSL* as we would expect, just like what the Hilbert school does. Carnap also freely allows the introduction of nonfinitary syntactical notions and explicitly acknowledges that the significance of the consistency proof for language II "must not be over-estimated" and "it gives us no absolute certainty that contradictions in the object-language II cannot arise… since the proof is carried out in a syntax-language which has richer resources than language II" (Carnap 1937, 129). He is also aware of the fact that the definition of mathematical truth for a formal language S (which is equivalent to "analytic in S") that includes classical mathematics outstrip what is formalizable in that language, i.e., cannot be defined in the language itself. But is this interpretation of SIM as a proposal to understand mathematics rather than a justification in the foundational sense a stable position which can avoid Gödel's circularity argument for the content of mathematics? Or is it just an apparent solution that is a will-o'-the-wisp? I will try to argue that the latter is more likely.

[58] For a typical substantive interpretation, see (Friedman 1988, 2001).

1.3.3.3.2 Problems for Transfinite Syntax and the Limits of Tolerance

The first point to mention is that there seems to be some uncertainty about what Carnap exactly means by syntax. In the introduction of *LSL*, he defined pure syntax as "nothing more than combinatorial analysis, or, in other words, the geometry of finite, discrete, serial structures of a particular kind" (Carnap 1937, 7). This conception is very close to Hilbert's idea of finitary mathematics and we could take it in this context to be finitary combinatorics or primitive recursive arithmetic[59]. However, in his discussion about general syntax Carnap freely introduced indefinite terms (non-recursive ones) and impredicative terms into his syntax language, leaving the permissibility of such terms in his syntax language to be merely a question of choice.

To be sure, the possibility of a common syntax language as weak as PRA comes from Gödel's famous technique of "arithmetization of syntax" used in the proof of his incompleteness theorem, with which Carnap is very familiar.[60] Most of syntactical notions of a formal system such as "being a formula", "being a direct consequence of" and most importantly "provable from" and "consistency" can already be expressed inside the formal system, as long as the system includes PRA.[61] Had such a conception

[59] This is the interpretation that is adopted in (Friedman 1988).

[60] Carnap gave a detailed presentation of the idea and execution of arithmetizing syntax in §18-§23 of *LSL*.

[61] This possibility again can be used as an argument refuting certain radical pluralism which regards even basic arithmetic as syntactical conventions, open to choice. PRA, at least has to be considered to be in a certain absolute sense meaningful truth, if syntax is to be possible and if consistency is a meaningful problem at all. Gödel also expressed the view that it is exactly his objectivistic conception of mathematics and metamathematics in general that makes the idea of "arithmetization of syntax" possible: "How indeed could one think of expressing metamathematics in the mathematical systems themselves, if the latter are considered to consist of meaningless symbols which acquire some substitute meaning only through metamathematics?" (Wang 1974, 9)

of syntax language been adequate for establishing the admissibility (consistency) of the syntactical conventions for mathematics, though still not strictly true, Carnap's claim that mathematics is void of content would have been be to a large extent justified and "the prima facie content of mathematics would turn out for the most part to be an illusion" (Gödel 1953/9b, 357–58). In addition, the existence of a common syntax language sufficient for the discussion and comparison of different object-languages representing different philosophical parties would also make Carnap's idea to resolve epistemological controversies via the principle of tolerance hard to refute. Unfortunately (or rather, fortunately), this weak notion of syntax language cannot serve the purposes expected, the most important of which is the definition of "analyticity" for sentences in the object language. In order to arrive at an adequate definition of "analytic in S" for a certain object language S, i.e., each logical and mathematical sentence or its negation has to be analytic in S, Carnap needs to assume the notion of all properties whatsoever, not only the ones that are definable in S.[62] This requirement, although Carnap denied that we thus "arrive at a Platonic absolutism of ideas" (Carnap 1937, 114) by speaking about "all properties", still makes the syntax language go far beyond PRA.[63] The same goes for his so-called proof that the axiom of choice is analytic, only by admitting the principle itself in the syntax language.

[62] It's interesting to note that Carnap arrived at this correct definition only with the help of Gödel, after Gödel pointed out a mistake in his originally proposed definition, see Gödel's letter to Carnap on 11, September 1932 in (Gödel 2003a, 347). Carnap's (or Gödel-Carnap's definition) can be regarded as a notational variant of the more familiar set-theoretic definition of truth by Tarski, see (Coffa 1987) for an interesting discussion between Carnap's and Tarski's definition.

[63] Carnap's denial of Platonism might also come from Gödel's indications, when Gödel stated in the footnote of the same letter (cf. footnote 42) (regarding the insufficiency of Carnap's original definition) that "This doesn't necessarily involve a Platonistic standpoint, for I assert only that this definition (for analyticity) be carried out within a definite language in which one already has the concept 'set' and 'relation'". The difference between Gödel and Carnap here is that for Gödel the definition of analyticity requires only other abstract concepts while Carnap resorts simply to a choice of complex syntax language.

Carnap, fully aware of this fact, seems not too troubled by it since "they only show that our definition of 'analytic' effects on this point what it is intended to effect, namely, the characterization of a sentence as analytic if, in material interpretation, it is regarded as logically valid" (ibid. 124) and what counts as logically valid itself is only a matter of choice based on reasons like expedience.[64] However, it is hard to insist or even to formulate the thesis of pluralism and the principle of tolerance, if, besides syntactical rules for the object language, we also allow the principle of tolerance in the syntax language, which seems inevitably to result in an infinite regress and Gödel's criticism will be justified. If mathematics is always needed again in showing that certain other mathematics comes from the syntactical or combinatorial properties of symbols, then this idea, SIM, even if right, can never be made in a transparent way. Is such a kind of regress vicious? Some think not, as Goldfarb and Ricketts would argue:

> To be sure, there is a regress, but it is not obviously circular or vicious unless one thinks that some foundational work must be done by the syntactical description of a language. If no such task is at issue, then the upshot is simply that we can never make the conventional nature of

[64] It's worth mentioning Gödel's view about the significance of the set-theoretic definition of truth as well in his next letter to Carnap. Fully aware of the fact that "analytic" (or "true") cannot be defined in the same system due to the simple truth-paradoxes, Gödel describes the interest of the definition as "not a clarification of the concept 'analytic', since one employs in it the concepts 'arbitrary sets', etc., which are just as problematic. Rather I formulate it only for the following reason: with its help one can show that undecidable sentences become decidable in systems which ascend farther in the sequence of types" (Gödel 2003a, 357). The word "problematic" to describe arbitrary sets should not mislead us into thinking that Gödel was skeptical about the concept of sets as he later claimed to be and that his avowed Platonism might have not been as strong as can be gathered from, for example (Gödel 1944, 1947, 1964). For other evidences from the 1930s and the general doubt, see (Davis 2005). The reasons against such misunderstandings are (1) "problematic" could just mean we haven't had a clear perception of the concept set, not that it doesn't exist or could never been made precise, (2) the basis of Gödel's realistic view of the concept set, i.e., the iterative conception, only begins to emerge at the time Gödel wrote the letter (1932), see for example (Zermelo 1930).

> mathematics fully explicit in any framework. The structure of Carnap's view is then coherent. Given the distinction between issues within a linguistic framework and issues between linguistic frameworks—a distinction that is always central to Carnap's thought—then the position is not circular so much as self-supporting at each level. If the mathematical part of a framework is analytic, then it's analytic; and so invoking mathematical truths at the level of the metalanguage is perfectly acceptable, since they follow from the adoption of the metalanguage. (Goldfarb and Ricketts 1992, 71)

According to these commentators, the cost to be paid for this lack of explicitness—i.e., the indispensable use of a metalanguage mathematically more expressive than the object language in order to define the consequence relation of the object language will not allow the conventional or non-factual nature of mathematics to be fully and explicitly displayed—can be compensated by foreswearing Carnap's version of conventionalism to be a foundational one. The foundation of mathematics, understood in an orthodox way, is expressed clearly by Gödel as consisting of two aims:

> At first these methods of proof have to be reduced to a minimum number of axioms and primitive rules of inference, which have to be stated as

> precisely as possible, and then secondly a justification in some sense or other has to be sought for these axioms, i.e., a theoretical foundation of the fact that they lead to results agreeing with each other and with empirical facts. (Gödel 1933c, 45)

It's obvious then that Carnap's position, abandoning the central notion of justification, will appear to be trivial or empty for Gödel, but since Carnap didn't aim at an epistemological reduction of mathematics to syntax or convention in the first place, Gödel's criticism doesn't apply to him either. There seems to be an inevitable standoff here.[65]

There are several problems with this interpretation however. Firstly, there is the minor interpretive problem of whether at the time of *LSL* Carnap was already so firm about the idea of internal framework problems and external problems between different frameworks, which appeared explicitly only a decade later[66] and which Goldfarb and Ricketts claimed to be "a distinction that is always central to Carnap's thought."[67] This is important as we mentioned in 2.2 that

[65] See (Goldfarb and Ricketts 1992, 71; Goldfarb 1995, 330) for such a strong reading.

[66] I.e., in (Carnap 1950a).

[67] That there might be a bigger difference in Carnap's view on the nature of foundational problems during his intellectual development than Goldfarb and Ricketts assume is exhibited in Carnap's different articulations of the justification of impredicative definition. In § 44 of *LSL* where Carnap discussed the question of the admissibility of impredicative terms, his answer is that "The proper way of framing the question is not 'Are indefinite (or impredicative) symbols admissible?' for, since there are no morals in logic, what meaning can 'admissible' have here? The problem can only be expressed in this way: "How shall we construct a particular language? Shall we admit symbols of this kind or not? And what are the consequences of either procedure?" It is therefore a question of choosing a form of language". (Carnap 1937, 164) While just a few years earlier, in his discussion of logicism, he deemed the problem of impredicative definition to be one of "the most difficult problems confronting contemporary studies in the foundations of mathematics" (Carnap 1931, 45). He considered Ramsey's justification of impredicative definition by presupposing a totality of properties already existing before their characterization by definition "theological mathematics", in contrast to intuitionistic "anthropological mathematics" (ibid. 49). Carnap then went on to provide his own method for the admissibility of impredicative definition by renouncing interpreting universal quantifier as infinite conjunction. It is beyond any doubt that Carnap was then not treating the question as

Gödel was more concerned with the conceptual problem of SIM rather than the possible peculiarities of Carnap's later philosophy. Secondly, it is hard to say why the necessary introduction of strong, transfinite, abstract syntax language for the syntactical program would show simply that we can never make the conventional nature of mathematics fully explicit in any framework rather than the conclusion Gödel gets, i.e., the non-feasibility of SIM. It is hard to imagine that we cannot, even in principle, manifest something which is purely (merely) our convention, in any single framework. The truth, I believe, is much more likely to be the opposite, i.e., we should be able to, at least in principle, makes our conventions totally apparent in a certain way.[68] A more serious problem for the allowance of strong syntax language and, as a consequence, Carnap's philosophy in general, is that the tolerance principle and consequent pluralism out of choices are not as innocent or neutral as they look. After abandoning the idea of a common syntax language like PRA, the various different choices of syntactical rules which are made possible by the so-called principle of tolerance cannot after all be treated on an equal basis as the principle should demand, since the standpoint adopted in the investigation of some object languages presupposes concepts or linguistic resources that those who favor simpler languages (or other similar frameworks)

---

only a choice of language forms.

[68] This idea is similar to an argument Gödel made in his Gibbs lecture against the view that mathematics is only our own creation, "for the creator necessarily knows all properties of his creatures, because they can't have any others except those he has given to them" (Gödel 1951, 311) and then the existence of absolutely undecidable propositions (one of the two disjunctions of the philosophical implications of the incompleteness theorem) would refute this view in favor of a realistic one admitting the objective existence of mathematical objects and facts. To the obvious objection that a creator need not necessarily know every property of what he constructs just like we may build a machine whose behavior we cannot predict in every detail, Gödel's reply to this "poor" objection is that "we don't create the machine out of nothing, but build them out of some given material" (ibid. 312).

rule out. There is a certain definite sense in which proponents of simpler languages cannot even understand the richer languages. Containing the basic PRA in their syntax language, the simple languages can manipulate the purely formal part of a richer language without, however, being able to understand it, i.e., to give it a meaning. This is exactly the criticism that Beth brings up against Carnap, the basic idea of which is that "the syntactical properties [like consistency] of a given object language O [depends] on the choice of a syntax language S"[69] (Beth 1963, 478). Some syntax languages, being so strong that they cannot be exhausted by deductive procedures, might be understood in a model-theoretically nonstandard way[70], unless they are "interpreted by reference to a certain presupposed intuitive model M" (ibid.). These results thus show the existence of "a limitation regarding the Principle of Tolerance" (ibid.), i.e., one can tolerate the other while not vice versa.[71] A fourth objection comes from Crocco (Crocco 2003), who argues that "Gödel did not overlook the novelty of Carnap's solution, and did not criticize him from an old-fashioned conception of science" (ibid. 21). In contrast with Gödel's foundationism, Carnap's position could be better characterized as an explicationistic one: SIM is not intended as a foundation or

[69] An example might be: consider first order Peano arithmetic PA, since Con (consistency of this system) is an undecidable proposition, we could use either S1 (PA+Con) or S2 (PA+¬Con) as a syntax language (in the strict sense of syntax), then the syntactical problem of whether PA is consistent obviously depends on the choice of a syntax language.

[70] Beth called it a "variant of the Skolem-Löwenheim paradox" (Beth 1963, 487). We don't need to go into the technical details, suffice it to say that the adoption of a richer syntax language than PRA does have unpleasant consequences for Carnap.

[71] Similar criticism has also been raised by Kleene in his review of *LSL*: "In general, it is not uniformly clear that, under translation of syntactical terms, a question will become more exact or the difficult will disappear by the principle of tolerance. There is a tendency for the inexactness or the difficulty to reappear, metamorphosed, in the syntax-language, either in the terms used to express the formal rules of the object language, or in the questions involved in the choice of the particular object language for a given purpose". (Kleene 1939, 87)

justification of mathematical truths, but as an explication of the importance and the use of mathematical concepts in science, while explication is used in the sense that Carnap used it in (Carnap 1950b), i.e., a transformation or rather replacement of the pre-formal ordinary concepts into exact, precise ones. In the case of mathematics syntactical rules serve as the precise and rigorous *explicata* of the ordinary and intuitive notions such as mathematical truth and consequence (the *explicanda*). Along with this conception of explication is the tolerance principle, which means we can use anything we want in the explication—in particular abstract and transfinite concepts and rules may also be allowed, without any ontological commitment to the objects to which these concepts are supposed to refer. However, as Crocco argues, even given this broad conception of SIM, consistency proof is still needed, since any explication of mathematics should not destroy our trust in the predictive power of mathematics, for which however, a consistency proof is necessary. That is to say, even if nonfinitary syntax and transfinite concepts or proof procedures are allowed in arriving at the same mathematical sentences as by using mathematical intuition,

> Mathematical intuition in addition produces the conviction that if these sentences express observable facts and were obtained by applying mathematics to verified physical laws (or if they express ascertainable mathematical facts), then these facts will be brought out by observation (or computation). Therefore syntax, if it is to be an acceptable substitute for

> mathematical intuition, must also yield sufficient reason for this expectation, and for this purpose a consistency proof is necessary. (Gödel 1953/9a, 340)

That is to say, even this explicationistic version of SIM is to be refuted: it cannot provide the same trust we have in the predictive power of mathematics, especially in applied mathematics[72] by using mathematical intuition. The last but not least objection for the interpretation of SIM as a proposal rather than a thesis and Carnap's general principle of tolerance and pluralism is that, if taken as a proposal, it has never been taken in the actual practice of either logic, mathematics, or philosophy. As Koellner put it,

> I do not think that he has said anything persuasive in favor of embracing a thorough radical pluralism as the "most expedient" of the options. The trouble with Carnap's entire approach (as I see it) is that the question of pluralism has been detached from actual developments in mathematics. To be swayed from the default position something of greater substance is required. (Koellner 2009b, 92)

[72] Gödel used two examples to show the necessity of consistency proof for syntactical conventions. One is the prediction of a computing machine which is known empirically to work reliably decomposing an arbitrary even number into two prime numbers and the other is the prediction (on the basis of empirically known physical laws such as the elasticity theory) that a bridge constructed in a certain manner will not break under a certain load. The parts of the bridge of the second example plays the same role as the elements (single electronic tubes) of the computing machine. In order for SIM to have the same prediction as a proof of the truth of Goldbach's conjecture or the truth of mathematics involved in the construction of a bridge, the consistency of those syntactical rules must be proven. Otherwise from the mere fact that Goldbach's conjecture follows from some arbitrarily assumed rules for handling symbols nothing whatsoever can be concluded about the result the machine will yield. See (Gödel 1953/9a, 339–40).

To this, I cannot agree more. Take SIM for example.[73] The proposal to view mathematics as syntactical rules will suffer the unpleasant consequence that the conventional or non-factual nature of mathematics can never be fully and explicitly displayed due to the necessary introduction of (at least in some respect) richer mathematics in the syntax of those syntactical rules. Rather than biting this Carnapian bullet we could have a much better and positive way out, i.e., accepting the fact of the inexhaustibility of mathematical contents and facts. For every syntactical system (considered as an effective axiomatic system) that attempts to capturing the content of mathematics, there will always exist a new and independent fact (within the realm of elementary arithmetic) outside of this system—such as the one stating the consistency of the system—for whose truth a new intuition or method of proof is indispensable. This, of course, is just the route that Gödel took, i.e., by considering this fact to be the basic consequence revealed by his incompleteness theorem and which is shown more concretely in the development of set theory under the general scheme of looking for new axioms.

1.4 Conclusion

In summary, we should say that even though the syntactical interpretation of mathematics (SIM), represented mainly by Carnap, could, as a technical program, in

[73] Peter Koellner discussed other examples from physics and set theory (especially about the undecidable propositions) to show that even in the face of equivalence (empirical or mathematical), we do have theoretical reasons rather than purely pragmatic ones to choose between theories, i.e., different linguistic frameworks, thus showing the vacuity of radical pluralism. See (Koellner 2006, 2009a) for details.

some respect succeed, i.e., by assuming propositions of an infinite length or transfinite and abstract syntax language, the philosophical aims which constitute its original purpose and main interest cannot be realized. SIM is possible only if we presuppose the truth of mathematics and mathematics is void of content only if we make this claim trivial, i.e., define content in such a way that mathematics cannot have content. The failure of such attempts adds evidence to a confirmation of the view that mathematical objects and facts have an independent existence, just like physical ones, which can be perceived by mathematical intuition, although only in an indistinct and incomplete manner as shown be the inexhaustible nature of mathematical truth.

2. Gödel and Russell: Logic, Paradox and Realism

2.1 Introduction

"Russell's Mathematical Logic" (hereafter as RML)[74] (Gödel 1944) occupies a special place in Gödel's life and work. On 18 November 1942 Paul Schilpp wrote to Gödel inviting him to contribute a paper about Russell's logic to a volume in his Library of Living Philosophers, adding that Russell himself not only appreciated Gödel's participation but also considered Gödel to be "the scholar *par excellence* in this field (logic)" (Gödel 2003b, 217). The invitation came at a time when Gödel was frustrated with his attempt to extend his independence proof for the axiom of choice to one for the continuum hypothesis.[75] It is difficult to know whether Gödel would have continued his efforts otherwise. As it happened, Gödel apparently began to concentrate on the Russell essay and the study of Leibniz around the beginning of 1943,[76] essentially turning his research from mathematical logic to philosophy and cosmology. This wide-ranging essay marks the transition of Gödel's attention from the quest for definite mathematical results in logic to investigations of a more distinctly

[74] This paper was first published in 1944 as a contribution to *The Philosophy of Bertrand Russell* (P. A. Schilpp 1944) and was reprinted twice in 1964 and 1972, with only editorial changes in the text. As usual, I will use the version in the authoritative *Collected Works* of Gödel. (Gödel 1990)

[75] See (Wang 1981, 657). As for the possibility of extending the method to the problem of the independence of CH, "Gödel developed a distaste for the work and did not enjoy continuing it. In the first place, it seemed at that time he could do everything in twenty different ways and it was not visible which was better. In the second place, he was at that time more interested in philosophy."

[76] In an unsent letter answering the questionnaire from Burke Grandjean, Gödel wrote that "the greatest phil. infl. on me came from Leibniz which I studied (about) 1943-46". (Gödel 2003a, 449–50)

philosophical and historical character. As Parsons correctly points out, this paper "is notable as Gödel's first and most extended philosophical statement" and "perhaps the most robust defense of realism about mathematics and its objects since the paradoxes had come to the consciousness of the mathematical world after 1900" (Parsons 1990, 103-104).[77] Despite the apparently exclusive focus on Russell's work on logic, this paper, as Gödel himself describes, is more "a history of logic with special reference to the work of Russell" (Wang 1981, 657). Gödel uses Russell's work as a reference point to put together his own reflections on the nature and the fundamental underlying concepts and axioms of mathematical logic, against the background of the historical course from Leibniz to Frege, Peano, Russell, himself and beyond. Just as Weyl remarked, Gödel's paper "is the work of a pointillist: a delicate pattern of partly disconnected, partly interrelated, critical remarks and suggestions" (Weyl 1946b, 210). The structure and organization of RML seem too loose and scattered to find the principal thread connecting all parts apart from dividng them into small sections indicated by Gödel's own writing cues.[78] However, except for the inconclusive yet insightful remarks concerning the history and nature of logic in the beginning and concluding part, referring especially to Leibniz's ideas, the rest of the essay can be seen as a refutation of constructivistic[79] views of logic and mathematics (views that

[77] For an earlier defense of Platonism in mathematics, see (Bernays 1935b).

[78] As is done by C. Parson in his introductory note to RML in the collected works of Gödel, by dividing the paper into 8 sections. For a somewhat different analysis of the structure of the paper, see (Crocco 2012; Wang 1987 chapter 11).

[79] As Gödel emphasized in the addition of the reprint that the term "construtivistic" used here refers mainly to a strictly nominalistic kind of constructivism, i.e., denying the objective existence of classes and logical concepts. It is thus different from either "intuitionistic" or "constructive" foundational theories, because "both schools base their constructions on a mathematical intuition whose avoidance is exactly one of the principle aim of Russell's constructivism". (Gödel 1944, 119)

deny the objective existence of, for example, concepts and classes), embodied particularly in Russell's no-class theory[80], and at the same time a defence of the realistic attitude towards objects and concepts of mathematics and logic. Focusing on this line of thought, I will divide my discussion into three parts. First I will discuss an argument which is usually overlooked in this paper, i.e., Russell's use of the mathematics-physics analogy to justify his realistic attitude towards the logical objects and the validity of inductive justification for principles and axioms in the foundation of mathematics. The merit and difficulty of this argument will be discussed. In the second part we will turn our attention to arguably Russell's most well-known contribution to philosophy, i.e., the theory of definite descriptions which was described by Ramsey as a "paradigm of philosophy" (Ramsey 1929, 263). We will approach the problem, however, from a logical rather than a linguistic point of view and discuss in some detail one of Gödel's contributions to philosophy, his so-called the slingshot argument, which has not always received the due attention as it deserves. And finally we will come to a closer examination about Russell's most important contribution to logic, the theory of types, whose philosophical rather than mathematical significance will be discussed in a few separate parts, particularly problems concerning the vicious circle principle, ramified type theory and simple type theory.

[80] "No-class theory" is a technical device to enable us to paraphrase talk about classes as talk about propositional functions so that we could pretend without any loss as if classes do not exist, irrespective of the problem as to whether they really exist or not.

2.2 Mathematics-Physics Analogy Argument (MPAA)

2.2.1 The Ontological MPAA

Gödel begins his discussion by saying that "what strikes one as surprising in this field (concerning the analysis of the concepts and axioms underlying mathematical logic) is Russell's pronouncedly realistic attitude" (Gödel 1944, 120) and quotes one of his favorite sentences from Russell's *Introduction to Mathematical Philosophy:* "Logic is concerned with the real world just as truly as zoology, though with its more abstract and general features".[81] Earlier on in a lecture called "*The philosophical implications of mathematical logic*" (Russell 1913) Russell is even more straightforward:

> Logic and mathematics force us, then, to admit a kind of realism in the scholastic sense, that is to say, to admit that there is a world of universals and of truths which do not bear directly on such and such a particular existence. This world of universals must subsist, although it cannot exist in

[81] In a footnote Gödel mistakenly says that "the above quoted passage was left out in the later editions of the *Introduction*", which is all the more surprising considering Gödel's extreme meticulousness in his writing. There was an interesting story about this remark that I cannot help mentioning here. Russell once in his autobiography reported his impression of Gödel: "Gödel turned out to be an unadulterated Platonist, and apparently believed that an eternal "not" was laid up in heaven, where virtuous logicians might hope to meet it hereafter." (Russell 1968, 356). Gödel later commented in an unsent letter draft: "Concerning my 'unadulterated' Platonism, it is no more unadulterated than Russell's own … [Gödel then mentioned the quoted sentence above] At that time evidently Russell has met the 'not' even in *this* world, but later on under the influence of Wittgenstein he chose to overlook it". (Gödel 2003a, 317)

the same sense as that in which particular data exist. We have immediate knowledge of an indefinite number of propositions about universals: this is an ultimate fact, as ultimate as sensation is. (Russell 1913, 293)

However, as Gödel noted, Russell's realistic attitude had been gradually decreasing over the course of time,[82] arguably mainly under the influence of Wittgenstein and it was always stronger in theory than in practice. As Gödel showed in detail in the ensuing discussion in the 1944 paper that it is just Russell's refraining from a more decided realism towards the logical and mathematical objects such as classes and concepts to which are due the difficulties in Russell's logical works.[83]

2.2.2 The Epistemological MPAA

Apart from the above ontological analogy between logic, mathematics and natural science, Gödel mentions then the second epistemological aspect of the analogy between mathematics and natural science which is enlarged upon by Russell concerning the justification of the primitive propositions and axioms of logic. Gödel only says that "in one of his early writings", but it's very likely to be his article entitled

[82] Russell was definitely a realist towards logical and mathematical objects in 1903 when writing PoM. But the discovery of his famous paradox and the ensuing theory of descriptions and the more general idea of logical construction rather than postulation of entities enabled him to dispense with lots of things he took to be existent earlier, like classes, propositions and logical concepts. In 1910 and after that, he still takes individuals and atomic properties and relations, thus universals in general, to be necessary for logic. For a masterful survey of Russell's ontological development, see (Quine 1966).

[83] We will come back to this topic in the discussion of Russell's no-class theory below.

"*On 'Insolubia' and their Solutions by Symbolic Logic*", published in 1906. There Russell is denying the necessity of the self-evident and infallible character of the axioms of logic against Poincaré:

> The method of logistic is fundamentally the same as that of every other science. … logistic is exactly on a level with (say) astronomy, except that, in astronomy, verification is effected not by intuition but by the senses. The "primitive propositions" with which the deductions of logistic begin should, if possible, be evident to intuition; but that is not indispensable, nor is it, in any case, the whole reason for their acceptance. This reason is inductive, namely that, among their known consequences many appear to intuition to be true, none appear to intuition to be false, and those that appear to intuition to be true are not, so far as can be seen, deducible from any system of indemonstrable propositions inconsistent with the system in question. (Russell 1906a, 193)

Russell's motivation for this kind of inductive evidence and inductive justification of logical axioms mainly comes from the problem of the axiom of reducibility and axiom of infinity[84] that he has to introduce in order to deduce classical mathematics

[84] The axiom of reducibility says roughly that for any higher order predicates/relations there exist a co-extensive predicative one, (i.e., one whose values are elementary propositions, that is, truth-functions of a finite number of atomic propositions) while the axiom of infinity, in one of its versions, claims that there are an infinite number of individuals in the world, or at least in the logical system under consideration. They are essential in Russell's logic system because without them lots of theorems of analysis and even number theory cannot be established. See (Copi 1971 section 2.2 and 3.5) for a very readable introduction to these axioms. See (Quine 1941) for a clear presentation of the mathematical content of PM.

from his logical systems, but which, unlike the other axioms of his primitive propositions, lacks the self-evidence as is traditionally required of a logical axiom. Thus, Russell argues in *Principia Mathematica* (PM hereafter)*:*

> "that the axiom of reducibility is self-evident is a proposition which can hardly be maintained. But in fact self-evidence is never more than a part of the reason for accepting an axiom, and is never indispensable. The reason for accepting an axiom, as for accepting any other proposition, is always largely inductive, namely that many propositions which are nearly indubitable can be deduced from it, and that no equally plausible way is known by which these propositions could be true if the axiom were false, and nothing which is probably false can be deduced from it. (Whitehead and Russell 1927, 59)

But once this idea of inductive justification emerges for this particular axiom, he goes on much further to say that the reasons for accepting an axiom or any principles are "always largely inductive" and that

> [I]f the axiom is apparently self-evident, that only means, practically, that it is nearly indubitable; for things have been thought to be self-evident and have yet turned out to be false. And if the axiom itself is nearly indubitable, that merely adds to the inductive evidence derived from the fact that its consequences are nearly indubitable: it does not provide new

evidence of a radically different kind. Infallibility is never attainable." (Russell 1910, 251)

It seems to Russell that the justification of an axiom, in its logical sense, can only be inductive and what self-evidence and intuition, "being themselves a psychological property and … therefore subjective and variable" (Russell 1913, 293), do is only "add to the inductive evidence". He elaborates this idea in a systematic discussion of the "regressive method" of discovering the premises of mathematics by distinguishing two different senses of "premises", namely the "empirical premise", which is the proposition from which we are lead to believe the propositions in question and the "logical premise", which is some logically simpler proposition or propositions from which, by a valid deduction, the proposition in question can be obtained (Russell 1907, 272). To conflate these two senses is to confuse the epistemological and the logical order and is the source of mistakes that a simpler idea or proposition is always easier to apprehend. Actually in dealing with the principles of mathematics, the axioms are too simple to be easy, but their consequences are generally easier than they are and we are lead to believe the premises because we can see that their consequences are true, instead of the other way round. And "the inferring of premises from consequences is the essence of induction; thus the method in investigating the principles of mathematics is really an inductive method, and is substantially the same as the method of discovering general laws in any other science." (Ibid. 274)

Gödel also mentions that this idea of Russell's resembles that of Hilbert's "supplementing the data of mathematical intuition" by axioms not given in intuition. Similar, but not identical, because the dividing line between data and assumption lies in different places according to Russell and Hilbert. Gödel's assertion here needs to be taken with a little care to avoid confusion. For Hilbert the main aim is to secure classical mathematics using only finitarily intuitive mathematics. One of the most quoted passages representing his ideas is the following:

> To make it a universal requirement that each individual formula then be interpretable by itself is by no means reasonable; on the contrary, a theory by its very nature is such that we do not need to fall back upon intuition or meaning in the midst of some argument. What the physicist demands precisely of a theory is that particular propositions be derived from the laws of nature or hypotheses solely by inferences, hence on the basis of a pure formula game, without extraneous considerations being adduced. Only certain combinations and consequences of the physical laws can be checked by experiment--just as in my proof theory only the real propositions are directly capable of verification. (Hilbert 1927, 475)

The philosophical interpretations of Hilbert's ideas are difficult and it's not easy to delineate one particular position.[85] However, apart from the methodological similarity,

[85] See the chapter on Gödel and Hilbert for a detailed discussion.

Hilbert's idea is definitely different from Russell's. Unlike Russell who is trying to elicit the principles of mathematics using logical evidence, Hilbert is satisfied with justifying the whole of classical mathematics with a consistency proof using only finitary mathematics, thus suspending the question of epistemology of those higher parts. Furthermore, unlike Hilbert's, so to speak, one-way plan, Russell's epistemology is more of a "back-and-forth" structure: he uses logically and mathematically evident propositions to obtain the inductively justified fundamental principles first, and then uses those principles to deduce the whole classical mathematics.

Regarding the epistemological MPAA, Gödel, speaking in an approving way, says that "this view has been largely justified by subsequent developments, and it is to be expected that it will still be more so in the future." (Gödel 1944, 121). The general idea of the regressive method for seeking axioms in the foundation of mathematics has indeed been quite common, especially in the awake of paradoxes like Russell's which bring intuitive certainty into doubt. Zermelo, for example, wrote that in deciding the axioms of set theory

> there is at this point nothing left for us to do but to proceed in the opposite direction and, starting from set theory as it is historically given, to seek out the principles required for establishing the foundations of this mathematical discipline. In solving this problem we must, on the one hand, restrict these principles sufficiently to exclude all contradictions and, on the

other, take them sufficiently wide to retain all that is valuable in this theory. (Zermelo 1908, 200)

Gödel himself mentioned this principle too, although in a somewhat different context, as a possible criterion of acceptability of new axioms. In his paper "*What is Cantor's continuum problem?*" (Gödel 1947, 1964) discussing the problem of looking for new set-theoretic axioms for deciding unsolved set-theoretic problems, Gödel says that:

> However, even disregarding the intrinsic necessity of some new axioms, and even in case it has no intrinsic necessity at all, a probable decision about its truth is possible also in another way, namely, inductively by studying its "success". Success here means fruitfulness in consequences, in particular "verifiable" consequences, i.e., consequences demonstrable without the new axiom, whose proofs with the help of the new axiom, however, are considerably simpler and easier to discover, and make it possible to contract into one proof many different proofs.. … There might exist axioms so abundant in their verifiable consequences, shedding so much light upon a whole field, and yielding such powerful methods for solving problems (and even solving them constructively, as far as that is possible) that, no matter whether or not they are intrinsically necessary, they would have to be

accepted at least in the same sense as any well-established physical theory.

(Gödel 1964, 261)

Despite the resemblance of the basic idea of inductive justification, we should note that there are essential differences and some shared difficulties between Russell's view and Gödel's own, which are worth a closer look.

First of all, we should note here that Gödel is speaking about the inductive, probable justification of set-theoretic axioms as a second criterion of truth alongside with the first and more important one, namely, the intrinsic necessity of some axioms provided by mathematical intuition. Epistemology of axioms for Gödel is thus, using a phrase from Maddy, "two-tiered"(Maddy 1990, 33): some concepts and axioms are justified intrinsically by their intuitiveness and others are justified extrinsically by their consequences.[86] For Russell, however, there is in principle only one method for a decision of truth for logical and mathematical axioms: inductive justification. Intuitions, "being themselves a psychological property and therefore subjective and variable" (Russell 1913, 293), at best, only add to the inductive evidence and cannot constitute a different kind of criterion.

Secondly, as to the nature of the evident propositions that inductively justify the axioms, Russell mentions two conditions: (1) they have to be "nearly indubitable", and (2) no other equally plausible way is known by which these propositions could be true

[86] However I don't agree with Maddy in attributing the property of "simpler" and "more theoretical hypotheses" to these two differently justified axioms, at least this is not part of "Gödel's Platonistic epistemology" that Maddy is talking about. (Maddy 1990, 33) The axiom of choice, for example, is regarded as intrinsically true for the concept of set for Gödel, though it's by no means simple. See later in this section for further discussion.

if the proposed axiom were false. As for Gödel, the success or fruitfulness in consequences also consists mainly of two aspects: (a) lots of verifiable propositions can be deduced from the axioms needing justification, and (b) some other intrinsic virtues for the whole theory, such as simplifying proofs and yielding powerful methods for solving problems, etc. So actually Gödel's criterion of what he regards as evidence in the inductive justification is in a strict sense weaker than Russell's in that he counts the pragmatic elements of (b) also as "success", i.e., fruitful consequences. Besides, it also seems that Russell's conditions (1) and (2), even if taken together, still constitute part of Gödel's condition (a), since Gödel allows evidence deducible in other different, although more complicated ways besides the axioms in question. As an example, he mentions that the axioms for the systems of real numbers, rejected by the intuitionists, have to be regarded in this sense to be justified "to some extent, owing to the fact that analytical number theory frequently allows one to prove number-theoretical theorems which, in a more cumbersome way, can subsequently be verified by elementary methods". (Gödel 1964, 261) However, it's difficult to attribute to the axioms under consideration with such kind of verifiable consequences a higher degree of truth than merely an instrumental value, which the intuitionist could readily agree.

The above discussion unveils the third point that we want to elaborate, i.e., the general difficulty for the epistemological MPAA, not only for Russell and Gödel. More precisely, it is the discrepancy between the intuitive validity of the epistemological MPAA and the scarcity of its important practical applications. The underlying problem centers around the notion of evidence: what exactly corresponds to sense perception or

observation in the mathematical or logical case? A difference, rather than an analogy, seems to be salient here. Take Russell's condition for evidence and his axiom of reducibility for example. It can be shown that with the addition of the axiom of reducibility in Russell's logic system we can obtain the otherwise impossible theorems of analysis like the least upper bound theorem and the full induction principle.[87] However, those principles certainly cannot be regarded as "nearly indubitable", at least by Russell. Suppose, on the other hand, that we restrict the domain of evidence to such an arithmetic theory that includes the standard axioms for the successor function, recursive definitions for addition, multiplication and exponentiation, with the induction schema where the induction property is limited to formulas without unbounded quantifiers. Such a theory, even if not indubitable for everyone, definitely deserves to be a candidate for being arithmetic of the highest evidence. Let us call it RA.[88] Then it is a provable fact that RA is interpretable within Russell's logical system (ramified type theory with the axiom of infinity, to be precise) (Burgess and Hazen 1998). This means that the axiom of reducibility is after all not justifiable from arithmetic evidence, in a strict sense. In the preface to the second edition of PM, Russell says explicitly that "This axiom [axiom of reducibility] has a purely pragmatic justification: it leads to the desired results, and no others. But clearly it is not the sort of axiom with

[87] I use 'induction' here in the sense of mathematical induction, not in the sense of inductive justification. "Full induction principle" means that we don't need to restrict the induction property no matter how complex it is, as long as it can be expressed in the language. An induction property P is the property that we want to establish for all natural numbers in the induction schema: $[P(0) \wedge \forall x(P(x) \rightarrow P(x+1))] \rightarrow \forall n P(n)$.

[88] RA is strictly weaker than primitive recursive arithmetic (PRA) (because not all primitive recursive functions are in RA), which is considered by many to be the formal counterpart of Hilbert's finitary mathematics. See (Tait 1981) for example. We don't need to concern ourselves here about the difference between indubitable in Russell's sense and finitarily intuitive in Hilbert's sense, suffice it to say that RA does seem to be a good formalization of "nearly indubitable" evidence.

which we can rest content. On this subject, however, it cannot be said that a satisfactory solution is as yet obtained… Perhaps some further axiom, less objectionable than the axiom of reducibility, might give these results, but we have not succeeded in finding such an axiom". (Whitehead and Russell 1927 Introduction to the 2nd Edition, xiv) Whether Russell's suspicion concerns the particular axiom of reducibility or the general principle of inductive justification resulting from MPAA, we don't know. What we do know is that, the failure of what for Russell may be the most important application of this principle in logical and mathematical practice will no doubt cast a shadow on the credibility of it. In the case of Gödel, due to his more flexible conception of evidence, matters become more complicated. As for his condition (b), pragmatic virtues for a whole theory such as "shedding so much light upon a whole field, and yielding such powerful methods for solving problems", the most natural example coming to our mind might be the famous axiom of choice. However, Gödel himself sees the axiom of choice as intrinsically true due to the concept of "set", since "nothing can express better the meaning of the term 'class' than the axiom of classes and the axiom of choice"[89] (Gödel 1944, 139). As for (a), one of Gödel's favorite examples is the justification of large cardinal axioms in set theory, whose consequences include, *inter alia*, in a somewhat surprising way, propositions in the field of arithmetic such as Diophantine equations. It's a provable fact that with the addition of the new axioms some undecidable and unsolvable problems about the

[89] Gödel used the word 'class' rather than 'set' because his discussion was in the context of Russell, who used 'classes' as more or less synonymous with 'sets'. 'The axioms of classes' are the standard other axioms of Zermelo-Fraenkel set theory (ZF hereafter).

solution of certain Diophantine equation will become consequences of the axioms. However, these propositions, although they can be formulated in the language of number theory, are usually either of a logical nature or very complex in appearance that can hardly be said to be a "natural" mathematical proposition.[90] On the other hand, usually we can verify these consequences by computation only up to any integer but not for all integers, which is not up to the highest degree of verification. In the expanded 1964 version of his 1947 paper where he first discussed the criterion of truth for new set-theoretic axioms, Gödel seems to be more cautious about this second criterion, when he writes that "this criterion [fruitfulness in mathematics and possibly also in physics], however, though it may become decisive in the future, cannot yet be applied to the specifically set-theoretical axioms because very little is known about their consequences in other fields" (Gödel 1964, 269). A concrete case might be illuminating here to show the difficult of the notion of evidence in the general structure of MPAA. In the 1947 paper on Cantor's continuum problem Gödel urged the search of an independence proof of CH relative to ZF based on observations that CH might be false.[91] One of the arguments he gave is that there are some "highly implausible consequences" of CH. Unlike the other unexpected and implausible results in point set theory (such as, e.g., Peano's curves), which are just a lack of agreement between our geometrical concepts and set-theoretical ones, Gödel seems to consider these highly

[90] There are some so-called "natural undecidable propositions" such as the Goodstein theorem (Goodstein 1944) and other combinatorial propositions discovered by Harvey Friedman, but they seem to still fall short of the criterion of "naturalness" of practising mathematicians. The possibility, of course, is open that a certain natural proposition will be provable only with the help of some higher order axioms.

[91] By the time of his writing, Gödel has already achieved the consistency proof of CH relative to ZF, showing that the negation of CH is not deducible from ZF. (Gödel 1940) Thus, if CH is false, then CH must also not be deducible from ZF (assuming its soundness), resulting in the undecidability of CH from ZF.

implausible consequences to be against the set-theoretical intuition, i.e., set-theoretical evidence.[92] It turns out soon that Gödel's prediction is correct when Paul Cohen proved the independence of CH in 1963 (Cohen 1963). So this seems to be a good case where the principle of inductive justification and MPAA is usefully and correctly applied. Nevertheless, despite the correct conclusion, this particular argument of Gödel has been put into question by later set-theorists. In his introductory note to Gödel 1947/1964, Gregory Moore argues that "there appears to be little evidence that analysts and set theorists now regard as 'paradoxical' the kinds of thin sets cited by Gödel" (G. H. Moore 1990, 165). He then mentioned the opinions of Paul Cohen and Donald Martin in the same place in support of the same view, quoting Martin as saying that "while Gödel's intuitions should never be taken lightly, it is very hard to see that the situation [with CH] is different from that of Peano curves, and it is even hard for some of us to see why the examples Gödel cites are implausible at all" (Ibid.).

The conclusion to be drawn from the above discussion of the epistemological version of the MPAA for justifying fundamental mathematical/logical axioms is that, on the one hand, notwithstanding its simple structure and intuitive appeal, this argument can hardly be said to have had frequent applications in practice, especially in decisive cases; on the other hand, even in the successful cases of application, the degree of justification of axioms in the inductive argument relies heavily upon what is intuitive and evident, an independent account of which seems necessary to evaluate the

[92] Burgess suggests a different interpretation here when he argues that the implausibility of these consequences is based not on set-theoretical intuition, but only intuition of a heuristic type.(Burgess 2014) However, Burgess's conclusion is based again on his entire denial of the existence of mathematical, including set-theoretic intuition. For our argument here, it does no harm to suppose the more natural set-theoretical intuition against the heuristic interpretation.

final success of the original argument. Thus, eventually it appears that Gödel is in a better position in insisting on a two-tiered epistemology and giving priority to the truth criterion of intrinsic necessity, as against Russell, who, influenced by the paradoxes, holds a fallibilist view towards even the axioms of logic and mathematics and considers all justifications to be essentially inductive.

2.3 The Logic of "the": Gödel on Russell's Theory of Description

In this section we will discuss first the context of the problem of definite description, emphasizing its relation to logic, and then present Frege's famous solution and the consequence shown by Gödel's slingshot argument. Next Russell's own solution and his reasons supporting it will be presented, with a final part trying to offer verdict on the problem.

2.3.1 The Logical Nature of the Problem

Gödel views Russell's treatment of the definite article "the" as an interesting example of his analysis of a fundamental logical concept, rather than a problem of a linguistic nature or a problem for the philosophy of language, as it is usually treated in the current philosophical discussion. In this respect, the analysis of "the" is on a par

with the analysis of other more central logical concepts like the universal and existential quantifier. This point of view is largely justified also by Russell's own comment that "the subject of denoting is of very great importance, not only in logic and mathematics, but also in the theory of knowledge" (Russell 1905, 103). What Russell is referring to here as to the importance for the theory of knowledge is of course his famous distinction between knowledge by acquaintance and knowledge by description, which he repeats both at the beginning and the end of his first article dealing with this problem, "On Denoting" (Russell 1905, hereafter OD). What is less known and what perhaps plays a greater role in motivating his famous theory of description is its connection with problems of mathematics and logic, especially the paradoxes.[93] What seems more likely to be the real original situation is that in the course of his pursuit of the solution of the logical paradoxes Russell also recognized the relevance of the problem of denoting, the solution of which will lead to a solution of the paradoxes as well. The stress on its importance for the theory of knowledge is more like a spin-off, so to speak, of his new theory, which he incidentally finds to be of great use for its application. This observation is already alluded to in his puzzling remark that the no-class theory rather than the theory of types is the source of his solution for the set-theoretic paradoxes.[94] For Russell classified both his theory of description and his no-class theory as important applications of the same logical

[93] See (Potter 2000, 119-128) for a detailed account of Russell's views on denoting from PoM to OD.

[94] In (Russell 1906b), after rejecting the theory of types as a solution for the paradoxes, he famously proposed three possible solutions: the "zig-zag" theory, the limitation of size and the "no-class" theory. In a note added on 5th February 1906, Russell said "I now feel hardly any doubt that the no-class theory affords the complete solution of all the difficulties [the paradoxes] stated in the first section of this paper". (Russell 1906b, 164) See (Quine 1966, 660–61) for a rebuttal of Russell's claim.

principle, i.e., substitution of construction for inferences in logic and mathematics, though the case for classes may be perhaps the most important of all the applications of this principle. Similar views can also be found in his posthumously published papers dealing with logic during the period 1903 to 1905. In his illuminating introduction to the corresponding volume, Alasdair Urquhart uses Russell's correspondence during the period between *Principles of Mathematics* (PoM) and OD to demonstrate that the goal of the development of the theory of descriptions in OD was to find a way around the paradoxes. As he writes,

> Most of the very voluminous secondary literature on Russell's Theory of Descriptions discusses it in isolation from its setting in the enterprise of the logical derivation of mathematics; the resulting separation of the logical and mathematical aspects of denoting is foreign to Russell's own approach. (Urquhart 1994, "Introduction", xxxii.)

One eye-opening and yet witty passage typical of Russell is the following: "Alfred and I had a happy hour yesterday, when we thought the present King of France had solved the Contradiction; but it turned out finally that the royal intellect was not quite up to that standard" (Ibid. xxxiii.). The connection between the improper definite descriptions and pathological sets such as the set of all sets which don't belong to themselves is made clear here. No matter how much significance of his theory of definite descriptions Russell attached later to the theory of knowledge, there is no

doubt that the difficulty in its logical sense must be sought if a full solution for the problem is to be achieved. Russell's realistic attitude can also be seen here, for he considered this whole question of the interpretation of descriptions as a question of right and wrong, rather than just a matter of mere linguistic conventions. This is reflected in one of Russell's reasons for rejecting Frege's proposal for being artificial and not a solution of the real matter:

> Frege, who provides by definition some purely conventional denotation for the cases in which otherwise there would be none. Thus "the King of France," is to denote the null-class; "the only son of Mr. So-and-so" (who has a fine family of ten), is to denote the class of all his sons; and so on. But this procedure, though it may not lead to actual logical error, is plainly artificial, and does not give an exact analysis of the matter. (Russell 1905, 109)

But as Gödel points out, Russell's realistic attitude "always was stronger in theory than in practice. When he started on a concrete problem, the objects to be analyzed (e.g., the classes or propositions) soon for the most part turned into 'logical fictions'[95] (Gödel 1944, 121). Russell's treatment of "the" as an "incomplete symbol" is something belonging to the same constructivistic order of ideas as his no-class theory,

[95] Russell's use of fiction is no doubt different from its contemporary use of "fictionism" as a philosophy of mathematical objects and truth, which is perhaps closers to its ordinary use. "Logical fictions", in the sense of Russell, need not necessarily mean that these things do not exist, but only that we have no direct perception of them.

exactly opposite to his realistic attitude.

2.3.2 The Problem of "the", Frege's Solution and Gödel's Slingshot Argument

The problem, as Gödel presents it, concerns what the so-called descriptive phrases (i.e., phrases such as, "the author of Waverley" or "the King of England") denote or signify and what the meaning is of sentences in which they occur. Frege's answer, the one which seems to be the most natural, is that "the author of Waverly" signifies Walter Scott, and in general, a descriptive phrase denotes or signifies the object it describes. But this somewhat naive[96] view, under closer examination, leads to unexpected difficulties. Surprisingly the difficulty which Gödel considers here as the most serious one facing the naive theory is not what we would usually expect, namely, the case where there seems to be no denotation at all.[97] Rather the difficulty is that, under certain "apparently obvious axioms", it follows that the sentence "Scott is the author of Waverly" signifies the same thing as "Scott is Scott"; and this again leads "almost inevitably to the conclusion that all true sentences have the same signification (as well as false ones)" (Gödel 1944, 122). Of course, Frege himself actually drew this conclusion and regarded "the True" as the common signification or denotation of all

[96] I use the word "naive" in the same sense as it is in "naive set theory".
[97] At least this is the difficulty most readers of OD would naturally conjure up. Two of the three puzzles dealt with in OD are problems concerned with empty descriptions. However as analysis below will make it clear, even for Russell himself, this apparent difficulty, if it is one at all, is of minor importance for his own theory of description.

true propositions, although he meant it "in an almost metaphysical sense, reminding one somewhat of the Eleatic doctrine of the 'One'". (Ibid.) Since this collapsing argument (also known as the "slingshot argument"[98]), a kind of argument which collapses intensional distinctions on the basis of simple assumptions about significations, is widely used later in philosophical discussions of meaning and reference, modality and propositional attitudes, we need to have a closer look at it. Before we get into Gödel's own slingshot argument we will present two other more popular ones by Church and Davidson respectively, and then contrast them with Gödel's.[99]

Alonzo Church, in his *Introduction to Mathematical Logic* presents, in an informal way, the slingshot argument[100] (Church 1956, 24–25). The argument goes as follows: The sentence (1) "Sir Walter Scott is the author of Waverley" must have the same denotation as the sentence (2) "Sir Walter Scott is the man who wrote twenty-nine Waverley Novels altogether," since the name "the author of Waverley" is replaced by another name of the same person; the latter sentence, it is plausible to suppose, if it is not synonymous with (3) "The number, such that Sir Walter Scott is the man who wrote that many Waverley Novels together, is twenty-nine," is at least so nearly so as to ensure its having the same denotation; and from this last sentence in turn, replacing

[98] For a more general discussion about slingshot argument in the case of truth, see (Stoutland 2003).
[99] For a detailed survey and more comprehensive account of different styles of the slingshot argument, see (Neale 1995).
[100] Church in his 1943 review (Church 1943) of Carnap's work *Introduction to Semantics* (Carnap 1942) already presents a similar argument. However the argument is used for a totally different purpose, namely, to support Church's view that denotation of propositions must be its truth-value. His 1956 argument is presented, however, in the context of discussing Frege rather than Carnap.

the complete subject by another name of the same number, we obtain, as still having the same denotation, the sentence (4) “The number of counties in Utah is twenty-nine.”

It is to be noted, however, that in order for Church’s argument to go through, two extra principles are needed: (1) two sentences, even if different in their explicit form, have the same references as long as they are synonymous or nearly so, and (2) names or descriptions which have the same reference or denotation can be substituted for each other without changing the reference of the whole sentence. While (2) looks more plausible, condition (1), considering the difficulty of the criterion of synonymy, alone already makes his argument hardly a logically strict one.

Davidson, in his “True to the Facts” presents a similar argument, basing it on somewhat different principles, against the correspondence theory of truth. Rather than discussing the reference or denotation of sentences, he speaks of the fact that a sentence is supposed to correspond. His argument goes as follows:

> Let ‘s’ abbreviate some true sentence. Then surely the statement that s corresponds to the fact that s. But we may substitute for the second ‘s’ the logically equivalent ‘(the x such that x is identical with Diogenes and s) is identical with (the x such that x is identical with Diogenes)’. Applying the principle that we may substitute coextensive singular terms, we can substitute ‘t’ for ‘s’ in the last quoted sentence, provided ‘t’ is true. Finally,

reversing the first step we conclude that the statement that s corresponds to the fact that t, where 's' and 't' are any true sentences. (Davidson 1969, 753)

Formally the argument looks like this: (where [s] means the fact that s, and (ιx) means "the x such that"):

1. [s]

2. [(ιx)(x=d ∧ s) = (ιx)(x=d)]

3. [(ιx)(x=d ∧ t) = (ιx)(x=d)]

4. [t]

Thus, any two true sentence s and t corresponds to the same fact, [s], or [t]. The reference above from (2) to (3) depends on the principle that we can substitute co-referring terms, similar to the second requirement in Church's argument. (1) to (2), however, relies on the principle that logically equivalent sentences correspond to the same fact, as the sentences in the brackets are logically equivalent in the sense that they must be both true or both false at the same time. This principle of logical equivalence, rather than Church's synonymous, seems to be clearer and makes Davidson's argument more general. Nevertheless, it is still quite a contentious claim to say that all logically equivalent sentences refer to the same thing, especially when the structures of the sentences become far more complicated than those dealt with in propositional logic. Gödel's own slingshot argument, however, endorses a much weaker condition than logical equivalence—what Stephen Neale has termed "Gödelian

equivalence" (Neale 1995)—making his argument the most powerful of all three. It simply requires that

(1) 'F(a)' and the proposition 'a is the object which has the property F and is identical with a' means the same thing.

Now the other two conditions needed in order to "obtain a rigorous proof" (Gödel 1944, 122 footnote 5) for the conclusion that all true sentences have the same signification (as well as false ones) along with the assumption of the principle of compositionality, i.e., (PC) a composite expression containing constituents which have themselves a signification depends only on the signification of these constituents, are that

(2) every proposition "speaks about something", i.e., can be brought to the form F(a); (otherwise the argument only applies to atomic sentences, rather than sentences in general)

(3) for any two objects, there exists a true proposition of the form $a=a \wedge b = b$ or $a \neq b$.

Now it's not straightforward to obtain the rigorous proof that Gödel had in mind, even given all the conditions Gödel listed, which are supposed to be sufficient. Gabriella Crocco, for example, presented an argument (Crocco 2012, 232) as follows: (suppose Fa and Gb are two true sentences which speak about two different objects)

i) F(a);

ii) $a = (\iota x)(F(x) \wedge x=a)$; according to condition (1);

iii) a = (ιx)(F(x) ∧ x=a ∧ b=b); by (3) and (PC), supposing that the two predicates "=(ιx)(F(x) ∧ x=a)" and "=(ιx)(F(x) ∧ x=a ∧ b=b)" are co-denotative because they have the same extension;

iv) b = (ιy)(y=b ∧ a=a); from iii) by (2) and (3)

v) b = (ιy)(G(y) ∧ y=b); from iv) by (3) and (PC)

vi) G(b), from v) by (1) and (PC)

The above reconstruction of Gödel's intended argument, as far as it looks to me, seems very doubtful for the reason that (1) step iii) to iv) doesn't follow from the principles Crocco referred to, and (2) Gödel, in his presentation of the slingshot argument, nowhere mentions the condition that "two predicates are co-denotative because they have the same extension", which seems to occupy a central place in Crocco's argument, i.e., step iii) and v).[101] A much more plausible formulation of Gödel's argument goes as follows:[102]

i) Fa,

ii) a = (ιx)(F(x) ∧ x=a), from i) by (1),

iii) (ιx)(x=a ∧ x≠b) = (ιx)(F(x) ∧ x=a), by (PC), suppose (ιx)(x=a ∧ x≠b) refers to a,

iii) (ιx)(x=a ∧ x≠b) = a

iv) a≠b

[101] I think Crocco's reformulation of the proof is influenced by Gödel's later remark (see our discussion below) that the principle of extensionality does not apply to concepts may be a solution the the puzzle of all true propositions being co-denotative. However, that's totally independent from the present argument and Gödel nowhere in the relevant passages mentioned the problem of concepts.

[102] This is my adaptation of a similar argument proposed by Stephen Neale, see (Neale 1995, 789–90).

By the same token, we could prove that Gb has the same signification as a≠b, thus the same as Fa. Mutatis mutandis where "a≠b" is replaced by a=a ∧ b=b, an alternative in Gödel's condition (3).

Although the slingshot can be made rigorously valid and seem devastating for certain philosophical views about meaning and reference when formalized in the above way, its conclusion can be evaded by challenging one or more of its assumptions. Russell himself takes the route to deny that definite descriptions refer to anything, although his main objection to Frege's distinction of sense and reference does not seem to stem from the slingshot argument, as we will now see.

2.3.3 Russell's Objection to Frege and His Reasons in Favour of His Own View

In this section we will first discuss Russell's objection to Frege, then his own solution and the reasons for supporting his views.

2.3.3.1 Russell's Objection to Frege's Distinction of Sense and Reference in OD

In *OD* Russell argued that Frege's theory of sense and reference was "an

inextricable tangle, and seems to prove that the whole distinction of meaning and denotation has been wrongly conceived" (Russell 1905, 113). Ironically many readers found the argument itself even more knotty and elusive. There are basically two different arguments and we will follow them in the reverse order as Russell presented them as it makes it easier for our discussions.

The first argument concerns identity statements, which is exactly what motivates Frege in drawing his famous distinction. As Russell said,

> That the meaning is relevant when a denoting phrase occurs in a proposition is formally proved by the puzzle about the author of Waverley. The proposition "Scott was the author of Waverley" has a property not possessed by "Scott was Scott," namely the property that George IV wished to know whether it was true. Thus the two are not identical propositions; hence the meaning of "the author of Waverley" must be relevant as well as the denotation, if we adhere to the point of view to which this distinction belongs. (Ibid.)

No one with even the slightest knowledge of Frege can fail to be surprised by this argument from Russell, who was supposed to be the first philosopher who really understood Frege and made him a public figure to the whole English academic world. [103] For Frege, with his distinction of sense and reference in hand, he could easily give

[103] Through an appendix in Russell's Principles of Mathematics in 1903, entitled "The Logical and Arithmetical

the answer that in the context of a query by George IV,[104] the normal sense becomes the reference, thus avoiding the triviality problem, but only at the cost of leaving how and why the reference should change. This surprise suggests to us that the difficulties in Russell's mind are not what's in Frege's, but curiously Russell thought that Frege's theory was roughly the same as his, when he equated Frege's distinction between sense and reference with his own distinction between "a concept and what the concept denotes". The only difference which he mentions is that Frege's distinction was more sweeping, since he considered all proper names to have both sense and reference while for Russell, only names formed from concepts by means of *the* can be said to have meaning and other proper names only signify without meaning. We will see how this conflation causes the trouble for understanding Russell's arguments and whether there are genuine problems for both Frege and Russell beneath or apart from the mere verbal confusions. We now turn to have a brief look at Russell's second argument, which is an assemblage of controversies and debates.[105]

Russell's main point is that, granting the distinction of sense and reference, we will always fail to refer to the meaning of a denoting complex,

> The difficulty may be stated thus: The moment we put the complex in a proposition, the proposition is about the denotation; and if we make a proposition in which the subject is "the meaning of C", then the subject is

Doctrines of Frege". (Russell 1903)

[104] Or more generally, in an indirect context.

[105] There are lots of literature about this relatively condensed part in OD, see, for example, see (Kremer 1994; Makin 1995)for a discussion of the argument structure in OD. I only refer to the conclusion of some of the more important papers, to my best knowledge.

> the meaning (if any) of the denotation, which was not intended. This leads us to say that, when we distinguish meaning and denotation, we must be dealing with the meaning: the meaning has denotation and is a complex, and there is not something other than the meaning, which can be called the complex, and be said to have both meaning and denotation. (Ibid. 112)

Again, one might wonder why instead of the phrase "the meaning of C" we cannot use "the meaning of 'C'" to refer to what we want, namely, the meaning. This is exactly the way Frege adopted and what Church observed in his refutation of Russell's argument:

> Russell applied quotation marks to distinguish the sense of an expression from its denotation, but leaves himself without any notation for the expression itself; upon introduction of, say, a second kind of quotation mark to signalize names of expressions, Russell's objection to Frege completely vanish. (Church 1943, 302)

Similarly, Searle accused of Russell suffering from his misunderstanding and inconsistency in his argument against Frege:

> Their faults spring from an initial misstatement of Frege's position, combined with a persistent confusion between the notions of occurring as a

> part of a proposition (being a constituent of a proposition) and being referred to by a proposition. The combination of these two leads to what is in fact a denial of the very distinction Frege is trying to draw and it is only from this denial, not from the original thesis, that Russell's conclusions can be drawn. (Searle 1958, 141)

To start with, it is almost inconceivable that Russell would commit such elementary mistakes. Closer inspection shows that if we read Russell's arguments as against his earlier own notion of denoting concepts in PoM (or denoting complex) then all these frivolous criticism could be easily dispensed with, and the quick solutions they suggest would become doubtful under a view about proposition and meaning and denotation. The problem caused by denoting concepts for Russell is that while they are parts, i.e., constituents of the propositions under consideration, the propositions don't refer to them, but about something different, something denoted by the denoting concepts. This somewhat mysterious function of "shift of subject matter" is characteristic of denoting concepts. Using Russell's own vivid example:

> The fact that description is possible—that we are able, by the employment of concepts, to designate a thing which is not a concept—is due to a logical relation between some concepts and some terms, in virtue of which such concepts inherently and logically denote such terms. It is this sense of denoting which is here in question... A concept denotes when, if it

> occurs in a proposition, the proposition is not about a man: this is a concept which does not walk the streets, but lives in the shadowy limbo of the logic-books. What I met was a thing, not a concept, an actually man with a tailor and a bank-account or a public-house and a drunken wife. ... If we wish to speak of the concept, we have to indicate the fact by italics or inverted comma. (Russell 1903,§ 56)

It is not that Russell never considered the possibility of the existence of some entity like Fregean sense, but that even if they existed, we could never refer to them, and they only "lives in the shadowy limbo of the logic-books". Searle's arguments, as far as it aims to show that Russell's argument was invalid against Frege's theory, is acceptable, but fails to show the reason why Russell would reject this distinction (i.e., the distinction between occurring as a part and a constituent of a proposition, and being referred to by a proposition) in the first place. Russell's insistence that what occurs in a proposition be the same as what is referred to by the proposition is induced by his conviction of realism, namely that if our assertions and knowledge are to be relevant to the real world and are to be objective, real objects must occur in those propositions which we assert, know to be true. The same line of thought even led him to the awkward and unintuitive idea that Mont Blanc, the real mountain, must occur and must be a constituent of the proposition that Mont Blanc is over 4000 meters high (Letter to Frege on 12/12/1904, in Frege 1980, 169). Again the same realistic attitude towards the relationship between logic and language prevents Russell from adopting

the way Church suggested, i.e., by introducing another mark for the linguistic sign itself and differentiating it from the sign indicating its sense. However easy it may look to be, it's foreign to Russell's thought, for Russell "the relation of meaning and denotation is not merely linguistic through the phrase: there must be a logical relation involved, which we express by saying that the meaning denotes the denotation"[106] (Russell 1905, 111). Sense belongs to the realm of logic, and not linguistics, if we have to refer to it in terms of a complex including some linguistic phrases as constituent, then the relation between sense and reference degenerates into an accidental and no longer a logical one. In the same way objections for Russell's first argument by assuming indirect sense and indirect references in oblique contexts vanish, for the simple reason that a proposition, being an objective complex, cannot change its meaning just because the sentences and expressions used to indicate them are embedded in a different context.

However, all the above arguments, rather than showing the correctness of Russell's theory, only indicates the difficulties which either Frege's theory of sense and reference or Russell's theory of denoting concepts must face in order to satisfactorily solve those puzzles which motivate their theory, like identity and contradiction.

[106] This idea is reinforced by materials in his unpublished work: "Thus it is the meaning, not the name, which denotes the denotation; and denoting is a fact which concerns logic, not the theory of language or of naming." See 'On Meaning and Denotation' (Russell 1994, 318).

2.3.3.2 Russell's Own Solution and His Reasons Supporting It

Russell's famous solution, to keep the principle that whatever is referred to in a proposition also occurs in it on the one hand, and to avoid the puzzling conclusion brought about by the slingshot argument that all true sentences indicate the same thing on the other hand, is to deny that a descriptive phrase denotes the object described. More precisely, he takes the view that a descriptive phrase denotes nothing at all but has meaning only in context, i.e., an incomplete symbol: "This is the principle of the theory of denoting I wish to advocate: that denoting phrases never have any meaning in themselves, but that every proposition in whose verbal expression they occur has a meaning." (Russell 1905, 105) According to his theory, the sentence "the author of *Waverley* is Scotch" is defined to mean: "There exists exactly one entity who wrote *Waverley* and whoever wrote *Waverley* is Scotch." In this way, a sentence involving the phrase "the author of Waverley" does not (strictly speaking) assert anything about Scott (since it contains no constituent denoting Scott), but is only a roundabout way of asserting something about the concepts occurring in the descriptive phrase.

The reasons which Russell invoke in favour of his own analysis are threefold. First is the semantic one that a descriptive phrase may be meaningfully employed even if the object described does not exist. Second is the epistemological reason that one may very well understand a sentence containing a descriptive phrase without being acquainted with the object described, whereas for Russell it is a fundamental

epistemological principle that every proposition which we can understand must be composed wholly of constituents with which we are acquainted. And last Russell's analysis for definite descriptions has the methodological advantage of avoiding postulating unnecessary entities, which can be substituted by logical constructions out of known entities. The method of "postulating" whatever we want over its possible logical construction has the same advantages of, using Russell's famous metaphor, "theft over honest toil" (Russell 1919, 71), especially when some set of supposed entities has neat logical properties which turn out to be possible to be replaced by purely logical structures composed of entities which have not such neat properties.

The first and the second reasons alone do not seem to favour Russell's over Frege's theory, at least under the condition that we give Fregean sense a proper and consistent interpretation. The third one, at first sight, does bring a formal advantage for Russell in that we can avoid in a logical system any special axioms about the particle "the" (just like in Frege's system), i.e., the theorems about "the" is made explicit to follow from the definition of the meaning of sentences involving "the" and other axioms. Even Gödel admits that this is a formal respect in which we may give preference to Russell's theory, but subsists "only as long as one interprets definitions as mere typographical abbreviations, not as introducing names for objects described by the definitions, a feature which his common to Frege and Russell" (Gödel 1944, 123–24).

2.3.4 After all, Russell or Frege?

Gödel's verdict for the problem of definite descriptions, in connection with his remark above that Russell's theory, by eliminating the problem using a sort of contextual definition, solves it[107] only under a nominalistic rather than realistic view about definitions, is quite explicit and yet enigmatic:

> As to the question in the logical sense, I cannot help feeling that the problem raised by Frege's puzzling conclusion has only been evaded by Russell's theory of descriptions and that there is something behind it which is not yet completely understood. (Ibid. 123)

It is very unusual to find in the published works of Gödel, who is always so cautious and meticulous about the content, a remark that is so indefinite about its interpretation. Since Gödel himself didn't specify what this something is that is not yet completely understood and didn't even give any hints, we have to be more or less speculative here.

One possible explanation, given by Virdi (Virdi 2009, 235) is that even if we exclude definite descriptions from the primitive notations, the slingshot argument can still be reformulated in terms of set-abstraction operators where there is no question that they refer or not. This is just what Davidson's slingshot argument looks like when

[107] If by a solution for "the" we mean to provide an adequate set of axioms for it.

the iota operations are replaced by set abstracts:

1. s

2. $\{x: x=d \wedge s\} = \{x: x=d\}$

3. $\{x: x=d \wedge t\} = \{x: x=d\}$

4. t

Virdi then makes the claim that "too much weight has been placed on imaginary problems concerning the *iota* operator"(Ibid.) and Gödel was right to be hesitant because "excluding definite descriptions from the primitive notation just creates the illusion of a solution". (Ibid.) The problem for this explanation, as I see it, is that a set-abstraction operator does not differ essentially from the *iota* operator, at least for Russell, since we can interpret set abstraction as a particular kind of definite description in a way such that $\{x: F(x)\} \stackrel{\text{def}}{=} (\iota\alpha)[\forall y(y \in \alpha \leftrightarrow F(y))]$. In this way, the problem of definite descriptions come back again, in the guise of sets or classes.

The other more possible interpretation, I think, of Gödel's somewhat intriguing yet unusual remark is that Russell only shifted the problem a step back. After all what Russell's theory amounts to, and what the so-called contextual elimination of definite descriptions achieves can be seen as reducing to two steps, first the replacement of the definite description by an indefinite description with uniqueness added, and second, the contextual elimination of sentences containing the indefinite descriptions in favour of what amounted to existential generalizations. The problems concerning denoting

and sense only get hidden beneath the veil of existential quantifier and variables.

Cashing out this idea, it is actually formally provable that within a purely extensional system, allowing only the principle of substitutivity of genuine singular terms (excluding definite descriptions), given the Russellian contextual definition for definite descriptions, they behave no differently from singular terms formally, that is to say, the substitution of definite descriptions for singular terms is still valid and the slingshot argument valid too.

1. $c=(\iota x)Fx$ Premise,

2. Sc Premise,

3. $(\exists x)((\forall y)(Fy\leftrightarrow x=y)\wedge x=c)$ Russellian Definition,

4. $(\forall y)(Fy\leftrightarrow a=y)\wedge a=c)$ Assumption, suppose a is the object satisfying condition (3),

5. $a=c$ Conjunction Elimination

6. Sa PSST (Principle of Substitution of Singular Terms);

7. $(\forall y)(Fy\leftrightarrow a=y)$ Conjunction Elimination;

8. $(\forall y)(Fy\leftrightarrow a=y)\wedge Sa$ Conjunction Introduction

9. . $(\exists x)((\forall y)(Fy\leftrightarrow x=y)\wedge Sx)$ Existential Generalization; 8

10. $(\exists x)((\forall y)(Fy\leftrightarrow x=y)\wedge Sx)$ Existential Instantiation, 3,4,9

11. $S[(\iota x)Fx]$

In a commentary on Bernays' review of his article about Russell, (Bernays 1946) Gödel wrote two very interesting remarks concerning Bernays' criticism of Gödel's sort of a lack of clarity about sense and denotation in his article:

> (a) Das Prob. Der Beschreibung ist durch "Sinn" und "Bedeutung" in befriedigender Weise gelöst. (The problem of description is solved in a satisfactory way by "sense" and "denotation".)
>
> (b) Das Extents. Axiom gilt nicht für Begriffe. (The axiom of extensionality does not hold for concepts.) (Reprint E of Textual notes for Gödel 1944, in Gödel 1990, 321–22).

These two short remarks suggest that another possible elaboration for Gödel's above quoted enigmatic remark is to focus on the problem of concepts and its relation to its extension and other objects. Along these lines, and mainly by relying on Gödel's unpublished philosophical writings, Max-Phil VIII, Crocco has made it quite convincing that there is a right solution for Gödel for definite descriptions different from both Frege's and Russell's, but keeping some elements of both and some ideas from Leibniz. (Crocco 2012, 235–37) There Gödel proposed a third kind of entity besides concepts and objects to be the denotation of definite descriptions, i.e., "individual concepts", which are neither things nor concepts, but are related to both of them. Things are spatial-temporal, concepts are usually independent from space and

time and only exist in logical space, but an "individual concept" occupies a well-determined portion of the logical space. An "individual concept" is thus like sense in that it is a perspective/mode of presentation on a particular possible denotation, and also like denotation in that it is what the descriptive phrase refers to. Therefore it is a combination of both ("i*n der Mitte zwischen Sinn und Bedeutung*"). (Ibid. 236) In this way the substitutivity of co-referring terms in Frege's sense is blocked, not for Russell's reason that some of them don't refer, but for the reason that they refer to different things, since concept has become part of the denotation of a denoting phrase which does not obey the principle of extensionality. However, the caution Gödel expressed in the 1944 paper and the nature of the textual evidence (unpublished notes) suggest to us that this conclusion cannot be taken to be conclusive, being at best suggestive, especially since it concerns the objective existence and nature of concepts. We will turn to this problem again in our discussions about intensional paradox below.

2.4 Paradoxes and the Theory of Logical Types

In this section we will discuss the most important work of Russell in logic, namely the theory of types envisaged by him to solve the paradox and found classical mathematics at the same time. We will first discuss the importance of the paradoxes, emphasizing its relation to the concept of logical intuition. Then we will separate two independent parts in Russell's original theory, according to either Ramsey's famous

distinction of logical and epistemological paradoxes or the different underlying principles for the different parts of the theory. Next we will come to Gödel's remarkable discussion of the so-called vicious circle principle since according to Russell it is the common source for all the paradoxes. The plausibility and difficulty in regard to this principle will be discussed. Russell's no-class theory and the ramified type theory, which seems to follow from the vicious circle principle will be treated, separating the philosophical and mathematical part. The final part will come to the discussion of simple type theory as a possible theory of concepts as independently existing objects and the accompanying intensional paradoxes.

2.4.1 Paradoxes and Logical Intuitions

The paradox of the set of all sets which don't belong to themselves, which bears Russell's own name[108], along with its final[109] famous solution known as "the theory of logical types"[110] are themselves a watershed in the history and development of modern logic. It combines all the previous results of particularly Frege and Peano and also provides a stepping stone for nearly all later important logical developments such as Gödel's incompleteness theorem, Tarski's definition of truth and Hilbert's program.[111]

[108] Independently discovered also by Zermelo, see (Rang and Thomas 1981).
[109] For a brief survey of the evolution of Russell's ideas about the solution of the paradoxes leading up to the theory of types, see (Urquhart 1988).
[110] First introduced in a crude form by Russell in the appendix of PoM in 1903, formally and systematically published in 1908 as "Mathematical logic based on the theory of types"(Russell 1908), crystallized in PM (1910-1913) and modified in the second edition (1925-1927).
[111] Both Gödel and Tarski's results are achieved in logical systems based on the theory of (simple) types. As for its relation to Hilbert's program, it should be acknowledged that Hilbert himself made a huge contribution to modern

It can without much dispute be said to be Russell's most important contribution of his investigations in the field of analysis of the fundamental concepts of formal logic. As Gödel put it,

> … the most important of Russell's investigations in the field of the analysis of the concepts of formal logic, namely those concerning the logical paradoxes and their solutions. By analyzing the paradoxes to which Cantor's set theory had led, he freed them from all mathematical technicalities, thus bringing to light the amazing fact that our logical intuitions (i.e., intuitions concerning such notions as: truth, concept, being, class, etc.,) are self-contradictory. (Gödel 1944, 124)

Here Gödel is obviously taking the paradoxes in a very serious sense; they reveal to him "the amazing fact that our logical intuitions are self-contradictory". This attitude towards the paradoxes is of course at complete variance with constructivists like Brouwer, who blame the paradoxes not on some transcendental logical intuition which deceives us, but on a gross error inadvertently committed in the passage from finite to infinite sets, or again in contrast with Poincaré, who ridicules the logicists' effort to reduce contentful mathematics to sterile logic, pointing out ironically that at least the occurrence of paradoxes shows after all that logic is not that sterile (Poincaré

---

logic in simplifying and modifying the other logical systems available to him and developed a system for his own purpose. However, it was PM that provided the first systematic formalization of nearly all mathematics and inspired and enabled Hilbert to take a more precise treatment of his ideas that proofs themselves can be regarded as mathematical objects, resulting again into one of the two basic elements. See (Zach 2003) for the role PM played in Hilbert's program.

1906, 1070). The view that Russell and Gödel shared of the importance of the logical paradoxes is a reflection of their realistic attitude towards logic, although the route they take to solve it diverges.

As for the nature of Russell's paradox, contrary to the general impression that it is of a set-theoretic, thus mathematical nature, Gödel speaks of it as "logical". Later in his paper on set theory, Gödel also remarked that set-theoretic paradoxes "are a very serious problem, not for mathematics, however, but rather for logic and epistemology" (Gödel 1964, 258). The apparent conflict can be readily reconciled by realizing that there are two totally different notions of set at work in our informal understanding of 'classes' or 'sets' here, a mathematical, set-theoretical one and a logical, philosophical one. The logical notion of set as concept-extension, embraced by Frege and Russell, is the idea that a set is obtained by dividing the totality of all existing things into two categories, one falling within and another out of a particular concept. In addition to another principle which seems to follow from the meaning of the term 'concept' that every propositional function/every predicate symbol defines a concept, we get the Russell paradox easily. On the other hand, as far as the concept of set needed in mathematics is concerned, the iterative one suffices, according to which a set is something obtainable from the integers (or some other well-defined objects) by iterative application (including transfinite iteration) of the operation "set of", such as the set of rational numbers (i.e., of pairs of integers) or of real numbers (i.e., of sets of rational numbers). Thus, Gödel remarks that the iterative conception of set "has never led to any antinomy whatsoever; that is, the perfectly 'naïve' and uncritical working

with this concept of sets has so far proved completely self-consistent" (Gödel 1964, 259). Gödel further elaborates this point in his conversation with Hao Wang. When asked about the apparent discrepancy, Gödel said,

> 8.5.4 The difference in emphasis is due to a difference in the subject matter, because the whole paper on Russell is concerned with logic rather than mathematics. The full concept of class (truth, concept, being, etc.) is not used in mathematics, and the iterative concept, which is sufficient for mathematics, may or may not be the full concept of class. Therefore, the difficulties in these logical concepts do not contradict the fact that we have a satisfactory foundation of mathematics in terms of the iterative concept of set. (Wang 1996, 270)

Even as serious paradoxes for logic and concepts, they don't necessarily force upon us the denial of the objective existence of concepts or classes (in the logical sense), as Russell did when he took the course of treating both classes and concepts as nonexistent and trying to replace them by our constructions. Quite the contrary, just like in the case of sets where the iterative conception of set finally gives us a correct picture of understanding the concept of set (in the mathematical sense), we can expect something similar in the case of classes and concepts, i.e., gaining a clear and distinct perception of those concepts by solving the paradoxes in a right way and maintaining our original logical intuitions to a large extent. Although Gödel was only talking about

logical intuition in general and didn't attribute the same epistemological weight as he did later in his 1964 set-theoretical paper, the analogy between sensory perception and abstract intuition of concepts leaps to the eye. Just like we could have illusions in the case of sensory perception, the logical paradoxes are just symptoms of where our logical intuition goes astray. Gödel compares the paradoxes to sense illusions and thinks that

> 7.4.3 The set-theoretic paradoxes are hardly any more troublesome for the objectivistic view of concepts than deceptions of the sense are for the objectivistic view of the physical world. The iterative concept of set, which is nothing but the clarification of the naïve—or simply the correct—concept of set, resolves these extensional paradoxes exactly as physics resolves the optical paradoxes by the laws of optics. (Wang 1996, 238)

Using unresolved logical (intensional) paradoxes to argue that concepts are unreal is just as fallacious as to argue that the outer world does not exist because there are sense deceptions. Surprisingly more can be shown by the paradoxes, the occurrence of which not only doesn't pose a threat for an objective existence of concepts, but can be used to prove a realistic view of the objective existence of concepts in exhibiting the non-arbitrariness of the way concepts can be formed even by correct principles. We will come back to this point in more detail in the last section on simple type theory and intensional paradoxes.

2.4.2 Separating the Two Type Theories

Russell's original theory for solving the paradoxes is referred to by the general name "the theory of logical types", which in PM actually consists of two different and distinct parts, i.e., what we would call simple type theory (STT) and ramified type theory (RTT) now.[112] It's Ramsey who first proposed this separation, as a consequence of a distinction between different kinds of paradoxes that need to be dealt with, which he attributed back to Peano. One kind of paradox (for example, Russell's paradox) consists of contradictions of a logical and mathematical nature since they involve only logical or mathematical terms such as class and numbers, while another kind of paradoxes (such as the liar paradox) is not purely logical and cannot be formulated in logical terms alone for the reason that they "all contain some reference to thought, language, or symbolism, which hare not formal but empirical terms"[113] (Ramsey 1925, 20–21). Based on this distinction, Ramsey then regarded as a defect of PM the lack of a separate solution for these two kinds of paradoxes: "These contradictions it was proposed to remove by what is called the Theory of Types, which consists really of

[112] It's impossible to give a full formal description of STT and RTT here, see (Copi 1971) for a classical introduction of both its philosophical and mathematical elements. The basic idea of STT is that objects of thought (or, in a nominalistic interpretation, the symbolic expressions) are divided into types, i.e., individuals, properties of individuals (or their extensions, sets), properties of properties of individuals, etc., and that a sentence of the form "a has a property F" is meaningful only if they are of the appropriate type, (F must be exactly one type higher than a, according to Russell). RTT then differentiates properties of the same type into different orders, according to the range of the bound variables occurring in their definition and new properties/propositional functions defined in terms of a totality of other properties must itself be of a higher order, such as the property "having all the traits to be a good general" must itself of a higher order than those traits referred to in its definition, although they are of the same type, i.e., both having individual persons as their argument.

[113] Peano called them "linguistic", and Ramsey preferred the term "epistemological", while they are more often referred to as "semantic" paradoxes nowadays.

two distinct parts directed respectively against the two groups of contradictions. These two parts were unified by being both deduced in a rather sloppy way from the 'vicious-circle principle', but it seems to me essential to consider them separately" (Ibid. 24). Whether the two kinds of paradoxes are really so different and whether one common solution is possible or not,[114] there are other independent reasons to separate STT from RTT. Ramsey already pointed out a second one: "…the essential distinction between order and type. The type of a function is a real characteristic of it depending on the arguments it can take; but the order of a proposition or function is not a real characteristic, but what Peano called a pseudo-function. … Order is only a characteristic of a particular symbol which is an instance of the proposition or function" (Ibid. 47). It's clear that Ramsey here is taking a realistic view towards STT which he believes to describe the real situation of objects, classes and concepts and which he takes it to be "unquestionably correct" (Ibid. 24), while RTT deals only with linguistic expressions and doesn't constitute a genuine logical theory, nor is it necessary. Gödel agrees with Ramsey on the need for separating STT from RTT when he writes that "[STT] in PM is combined with the theory of orders (giving as a result the 'ramified hierarchy') but is entirely independent of it and has nothing to do with the vicious circle principle" (Gödel 1944, 127), but bases this separation on a different reason, i.e., the underlying principles of STT and RTT are essentially distinct and cannot be unified in a rather "sloppy" way from the vicious circle principle. It is Gödel's central

[114] Russell himself never views the paradoxes as essentially of two kinds, possibly due to his conception of class as nonexistent entity: if class as an extensional entity is to be reduced or explained away as propositional functions of an intensional character, surely their solutions must be sought in the same way. For attempts to give a common solution to all the paradoxes, based on an idea of Russell, see (Priest 1994).

criticism of Russell's method and ideas in PM that no separation is made at all between the general requirements of a solution of the paradoxes and those subsisting only for Russell's own constructivistic point of view, and it won't be surprising that the paradoxes would have different solutions according to the different interpretations of the central terms occurring. Certain methods may indeed block the occurrence of the paradoxes, but only under certain particular philosophical assumptions. This in turn might make the proposed solutions ad hoc, or even worse, a misunderstanding of the original paradox in its full sense and thus cannot be thought of as a genuine "solution" of the paradoxes. To take as an example (and to facilitate our discussions below) possible different interpretations of logical terms depending on either a realistic point of view or a constructivistic/nominalistic view, Gödel distinguishes 'concepts' and 'notions', defined by propositional functions as combinations of symbols. Concepts should be understood in the objective sense as "properties and relations of things existing independently of our definitions and constructions" (Ibid. 128), while a notion is just a symbol together with a contextual definition, i.e., a rule for translating a sentence containing the symbol into such ones as do not contain it, so that there is no need to assume a separate object denoted by the symbol, that's to say, it appears as a mere fiction, in Russell's sense.[115] Formally, any two different definitions of the form $\alpha(x) = \beta(x)$ can be assumed to define two different notions while a concept generally allows multiple descriptions. Take the definition of the number two, for example. As a

[115] One may doubt whether such a concept of notion is consistent or not in that we need to assume some primitive notions irreducible to any other to avoid the infinite regress. It doesn't rule out the possibility, however that all the abstract notions can be contextually defined by a few primitive ones.

concept, two may be understood as something under which fall all pairs and nothing else. On the other hand, there is certainly more than one notion satisfying the above condition, such as different definitions of the ordinal two in set theory.[116] That different diagnosis and solutions of the paradoxes become necessary depending on whether it's a question of concepts or notions, we will see below, especially in relation to the vicious circle principle, which Russell regards as the source of all paradoxes.

2.4.3 The Vicious Circle Principle (VCP) and Its Consequence

By analyzing the common assumptions leading to paradoxes, Russell soon came to the conclusion that the erroneous axiom consists in assuming that for every propositional function (or property) there exists a class of objects satisfying it, or every propositional function can be treated as a separate entity, separable from its argument and distinct from the combination of symbols expressing this function. There are some propositional functions whose extensions are, in a sense "self-productive" and thus can never be collected into a single unit, i.e., to form a class:

> The contradictions result from the fact that, according to current logical assumptions, there are what we call self-reproductive processes and classes. That is, there are some properties such that, given any class of terms all

[116] Thus, the so-called Benacerraf problem for identifying numbers (Benacerraf 1965) exists only for number two interpreted as a notion, but not a concept.

> having such a property, we can always define a new term also having the property in question. Hence we can never collect all the terms having the said property into a whole; because, whenever we hope we have them all, the collection which we have immediately proceeds to generate a new term also having the said property. (Russell 1906b, 144)

After abandoning the existence of a class or concept in general he refers to those of which a legitimate totality can be formed as "predicative" and seeks a general criterion for being predicative. And it's Poincaré's answer that motivates Russell's adoption of the VCP.[117] By analyzing Richard's paradox and the Burali-Forti one, Poincaré characterizes non-predicative definition as those that contain a vicious circle due to accepting the actual infinity: "The definitions which must be considered non-predicative are those which contain a vicious circle. A definition containing a vicious circle defines nothing" (Poincaré 1906, 1063). And later he offered an alternative characterization as follows:

> We draw a distinction between two types of classifications applicable to the elements of infinite collections: the predicative classifications, which cannot be disordered by the introduction of new elements; the non-predicative classifications in which the introduction of new elements necessitates constant modification. (Poincaré 1910, 1073)

[117] For more details about the debate between Russell and Poincaré, see (Goldfarb 1988; Detlefsen 1992).

In a reply article to Poincaré, Russell formulated his own version of VCP[118]:

> I recognize, however, that the clue to the paradoxes is to be found in the vicious-circle suggestion; I recognize further this element of truth in M. Poincaré's objection to totality, that whatever in any way concerns all of any or some (undetermined) of the members of a class must not be itself one of the members of that class. In Mr. Peano's language, the principle I wish to advocate may be stated: "Whatever involves an apparent variable must not be among the possible values of that variable". (Russell 1906a, 198)

In the mature work PM Russell gave several different formulations of VCP again, as if it doesn't matter which one to adopt. He first presents it as:

> (1) Given any set of objects such that, if we suppose the set to have a total, it will contain members which presuppose this total, then such a set cannot have a total. By saying that a set has no total, we mean, primarily, that no significant statement can be made about all its members. (Whitehead and Russell 1927, 37)

[118] Without any mention of the problem of actual infinity repudiated by Poincaré. Russell doesn't think infinity plays a vital role in the paradoxes, as can be shown in the liar paradox. In a typical witty way, Russell objects to Poincaré by asking "Has this man [the liar] forgotten that there is no actual infinite?" (Russell 1906a, 197)

Just a few sentences later, Russell speaks of the same principle as follows:

> (2) Whatever involves all of a collection must not be one of the collection; or conversely: 'If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total'. (Ibid.)

Synthesizing these formulations into one, we get "no totality can contain members definable only in terms of this totality, or members involving or presupposing this totality". This seemingly innocent formulation of VCP is really, according to Gödel, three principles corresponding to the phrases "only definable", "involving", and "presupposing", the latter two being more plausible than the first. However, it's the first form which is more interesting and important for the derivation of mathematics from logic, which is exactly what Russell tries to achieve, after Dedekind and Frege. Classical mathematics, however, demonstrably contains impredicative definition inhibited by this form of VCP. Consider the case for natural numbers. Frege famously defines the property of being a natural number as follows: n is a natural number if and only if n has all the hereditary properties of zero—where a property is hereditary if, whenever some number has it, so does its successor. The definition is not circular because he can define the successor relation without explicit reference to natural numbers. However, the property of being a natural number itself is one of the hereditary properties of zero. So Frege is defining the property of being a natural

number in terms of a quantification over a totality including that very property, which makes his definition impredicative.[119] In a similar way, the set-theoretic definition of the set of natural numbers as the intersection of all inductive sets is impredicative too because the set of natural numbers of itself one of the inductive sets in its definition.

As for Russell's own logical systems in PM for actually building up logic and then mathematics, in its first edition with the assumption of the axiom of reducibility, classical mathematics can indeed be obtained based on it, which however just shows that VCP in the first form is violated by Russell himself since classical mathematics include lots of impredicative definitions. In the second edition after abandoning the axiom of reducibility Russell introduced a new axiom of extensionality, i.e., propositional functions can occur in propositions extensionally only through their values irrespective of their order. This axiom then has the consequence that any propositional function can take as its argument any function of appropriate type, whose extension is defined, thus again violating the VCP for propositional functions, which requires that nothing defined in terms of a propositional function can itself be an argument for this function.[120] So in general, Russell's own work can be seen as a *reductio ad absurdum* argument for the claimed underlying principles VCP and the consequent impredicative definition. The fact that impredicative definition is necessary in classical mathematics can just as well be considered as an argument that VCP is

[119] The impredicativity of either Frege's or the usual set-theoretical formal definition of natural numbers does not, however, determine whether the concept of natural number itself is impredicative. Most interesting predicative systems of analysis are predicative given the natural numbers. See (Parsons 1992) for a view arguing for the impredicativity of the concept of natural numbers. Weyl gives a more common example of impredicative definition for the proof that any set of non-negative real numbers has a greatest lower bound, see (Weyl 1946a, 4–5).

[120] This follows from VCP in its first form plus the extra assumption that every propositional function presupposes the totality of its values, therefore also the totality of all its possible arguments.

false, rather than that classical mathematics is false, or at least problematic. What an impredicative definition of a concept essentially prohibits is the possibility of a total construction of its meaning in terms of other given concepts, i.e., a contextual definition for it is no longer possible since in translating a sentence with the defined notion we are forced back to a group of concepts which will contain the defined one again.[121] As to notions in the aforementioned constructivistic sense, it is indeed true that paradoxes can be generated with a vicious circle, for it is clear that a thing to be constructed certainly cannot belong to the totality of things from which it is to be constructed. However, if it's a question of independently existent concepts (or their extensional counterpart, classes), there is nothing in the least absurd in the existence of totalities containing members which can be uniquely defined by reference to this totality, such as the tallest man in a group. So the first form of VCP doesn't apply to either concepts or classes in the realistic interpretation, but only the third form of VCP (in the sense of 'presuppose') does, if 'presuppose' means "presuppose for the existence" and not "for the knowability". As for the second form of VCP (in the sense of 'involve'), Gödel thinks that it doesn't apply to classes understood as a certain kind of structures since a certain structure can involve a totality of which it is a part.[122] As for classes in the sense of pluralities or totalities, the second form of VCP looks very plausible in requiring that a class cannot involve a totality containing itself, but only other classes 'below' itself. The modern Zermelo-Fraenkel conception of sets, as

[121] At least this is the case if 'all' means an infinite conjunction.
[122] For example, the structure of the series of integers contain itself as a proper part and "it is easily seen that there exists also structures containing infinitely many different part, each containing the whole structure as a part". (Gödel 1944, 130)

forming a cumulative hierarchy based on iteration, can be seen as a realization of this idea. STT in the extensional interpretation can also be regarded as conforming to the second form of VCP, but with the unnecessary additional restrictions of excluding mixed types and dealing only with finite types.[123] Finally for concepts as existing objectively, the second form of VCP doesn't apply either. The fact that concepts (properties or relations) and propositions formed out of them sometimes will have to contain themselves as constituents of their content (or of their meaning) only makes it impossible to understand them in a constructive way (i.e., explain them by reference to other primitive entities), but in no way proves that these concepts have no meaning at all. Gödel mentions the example of his original famous undecidable sentence as one that "contains as parts of their meaning not themselves but their own formal demonstrability"[124] to show that certain circularity and self-reflexivity of impredicative properties, prohibited by the VCP, is not only harmless but rather helpful sometimes.

So, both with the need to accomplishing Russell's original logicist program and on its own accounts, we have good enough reasons to reject VCP in its first form and can put into better use of its second form as shown above. Russell, however, didn't take either of these two arguments due to his constructivistic tendency to build up logic as far as possible without assuming the objective existence of both classes and concepts embodied in his famous no-theory and RTT, to which we will turn our discussion below.

[123] Early in a lecture in 1933, Gödel already expressed the view that ZF is "nothing else but a natural generalization of the theory of types, or rather, it is what becomes of the theory of types if certain superfluous restrictions are removed".(Gödel 1933c, 46)

[124] With the additional requirement of soundness, formal demonstrability of a proposition is equivalent to itself.

2.4.4 Russell's No-class Theory and RTT

The device of "no-class theory" was first outlined by Russell as the third possible solution of dealing with the paradoxes in his 1906 paper (Russell 1906b), along with the method of zig-zag theory and the limitation of size. The latter two methods share the common feature of trying to find some criterion for classes and concepts to exist, without rejecting or acknowledging them in general.[125] Russell's preference for the no-class theory, no doubt, is based on the difficulty of setting up a restrictive condition required for the other two methods without artificiality and arbitrariness. According to his no-class theory, classes or concepts never exist, or at least never need to be taken as existent, and sentences containing these terms are meaningful only to such an extent as they can be paraphrased as a *façon de parler*, a manner of speaking about other things.[126] This is exactly the same strategy Russell found earlier in his theory of descriptions to overcome certain difficulties related. Unlike the case of descriptions, Russell appears to be more neutral about the existence of classes, as he wrote in PM that

[125] Gödel mentions Quine's system (Quine 1937) as a development in accord with the ideas of zig-zag theory, and axiomatic set theory as an elaboration of the limitation of size idea. However, it is to be noted (as also pointed out by Gödel) that the principle of the limitation of size appears only as a consequence, rather than as the basis for axiomatic set theory, thus we should avoiding talking about "the limitation of size" as the guiding idea of axiomatic set theory, as can be easily ignored. See (Giaquinto 2002 chapter VI) for a clarification of this point.

[126] See *20 of PM for details. An example will show the general idea. By defining $\Phi(\{x: G(x)\}) := \exists H[\forall x[G(x) \leftrightarrow H!(x)] \wedge \Phi(\lambda x. H!(x))]$, we can translate the sentence that "the class of Gs has the property Φ" into another one without mentioning classes at all: there exists a predicative property H, co-extensive with G, and H has the property Φ, (assuming the axiom of reducibility).

> In the case of descriptions, it was possible to prove that they are incomplete symbols. In the case of classes, we do not know of any equally definite proof, though arguments of more or less cogency can be elicited from the ancient problem of the one and the Many. It is not necessary for our purposes, however, to assert dogmatically that there are no such things as classes. It is only necessary for us to show that the incomplete symbols which we introduce as representatives of classes yield all the propositions for the sake of which classes which be thought essential. When this has been shown, the mere principle of economy of primitive ideas leads to the non-introduction of classes except as incomplete symbols. (Whitehead and Russell 1927, 72)

This passage is striking in a few aspects. The talk of "proof" again shows Russell's realistic attitude towards these problems, but we can easily doubt the degree of authenticity of the realistic position when he puts "the principle of the economy of primitive ideas" before the ontological question whether classes exist. Surely for a realist, the truth of the existence problem cannot be solved dogmatically as Russell claims, but neither could it be suspended in terms of an economy of ideas. For, if classes really exist, then a theory based on the assumption that they don't will sooner or later suffer all kinds of difficulties and prove themselves inadequate. And this is indeed what happened to Russell's logical system in PM. He did pursue the no-class theory in detail and reduced sentences containing classes to those without them so that

classes can be dispensed with, but only under the assumption that a concept exists whenever one wants to construct a class. But as to how concepts are to be reduced to notions, which is a necessary requirement by the original motivating ideas of no-class theory, it's not unquestionable at all.[127] Either Russell's own measurement of assuming the axiom of reducibility to enrich the basic data of the construction, or Ramsey's device of dealing with truth-functions with infinitely many arguments as a method of interpretation, on closer considerations, all prove to amount to the same thing as assuming the independent existence of classes or of concepts.[128] This essentially negative result[129] is only a refutation of the constructivistic view, and at the same time a confirmation of the correctness of the realistic position of assuming the objective existence of classes and concepts. Gödel thus writes

> The whole scheme of the no-class theory is of great interest as one of the few examples, carried out in detail, of the tendency to eliminate assumptions about the existence of objects outside the "data" and to replace them by constructions on the basis of these data [understood as logic without the assumption of the existence of classes and concepts]. … All this is only a

[127] Quine had a similar criticism to Russell's logical system by showing that Russell only reduces classes to propositional functions, but since these functions serve as variables of quantification and must be treated as real objects rather than linguistic formulas, the reduction is merely that of classes into properties or attributes (Quine 1941). However, Quine's objection is valid only with his own criterion of ontological commitment, which don't concern us here.

[128] Ramsey took our inability to form propositions of infinite length as a mere accident and considered infinite constructions to be possible, but this is, if at all, too high a criterion of construction to be of real practical use. As Gödel questioned: "what else is such an infinite truth-function but a special kind of an infinite extension (or structure) and even a more complicated one than a class, endowed in addition with a hypothetical meaning, which can be understood only by an infinite mind?" (Gödel 1944, 132)

[129] Gödel pointed out clearly that analysis cannot be obtained on the basis of PM (second edition), and discovered a mistake in Russell's so-called proof for the validity of the induction principle for the natural numbers, so the proof is "certainly not conclusive". We know very well now that Russell's proof is wrong and even arithmetic cannot be obtained in his system, due to Myhill. See (Myhill 1974).

> verification of the view that logic and mathematics (just as physics) are built up on axioms with a real content which cannot be "explained away". (Gödel 1944, 132)

This type of argument for a realistic attitude towards the objects of logic and mathematics shows one unique aspect of Gödel's realism, i.e., he always discusses it in the context of concrete problems and programs aimed at eliminating those objects and defends it based on definite mathematical and logical results, rather than purely by a priori philosophical arguments. In the case of Russell, the failure of no-class theory shows that, contrary to Russell's belief that no assumption about their existence is needed in mathematics, concepts and classes are necessary to obtain a satisfactory system of classical mathematics. Rather than the radical method of denying their existence at all, the more conservative way is to admit their objective existence and set up an axiomatic system by an analysis of these concepts according to the realistic view, as is exactly what happened in the actual historical development of mathematical logic.

On the other hand, RTT and the restrictions it brings along with the hierarchy of orders seem to be a natural consequence of Russell's constructivistic attitude and in particular the no-class theory, rather than just as an ad hoc device for avoiding the (semantic) paradoxes. That is to say, if indeed concepts and classes are just our logical constructions from primitive objects and don't belong to the "inventory" of the world, then the ramified hierarchy is necessary to guarantee the consistency of the process of the constructions: we are only allowed to go up to the next order with the condition

that the quantifiers in the higher order refer either to structures or objects in the original order or lower ones. As Gödel put it, "it is not as if the universe of things were divided into orders and then one were prohibited to speak of all orders; but, on the contrary, it is possible to speak of all existing things; only, classes and concepts are not among them; …they are introduced as a *façon de parler*" (ibid. 133). However, if we separate the basic ideas of RTT, a step-by-step logical construction from a basic system onwards, from the conception of it as an ontological theory (as is the case in Russell who denies the existence of classes and concepts apart from the elementary ones and individuals in the domain of logic) and from the conception of it as a foundation for building up mathematics, then RTT could prove to be a very useful mathematical method for logical inquiries. First, it can hardly be doubted that a predicative system based on the VCP principle is more elementary and enjoys a greater clarity than an impredicative one, as can be seen, for example, from the difficulty of consistency proofs of the theory under consideration.[130] Besides, the constructible model Gödel used in proving the consistency of the axiom of choice and Cantor's continuum hypothesis relative to ZF can be regarded as an application of the extension of the basic idea of RTT.[131] Moreover, the contemporary analysis of the notion of predicativity and the exploration of the extent and limit of predicative theories[132] in

[130] RTT can be proved to be consistency while impredicative systems like STT and ZF cannot. See (Fitch 1938).
[131] Gödel extends the hierarchy to arbitrarily high transfinite levels (including the non-constructive ordinals) in his constructible model and this again can be seen as the "cash value" of his Platonism, even in employing constructive methods. On the other hand, this gives us another reason to doubt the correctness of the constructible hypothesis (all sets are constructible) for "it is not a conceptually pure proposition because it allows ordinal numbers definable only by impredicative definitions or not definable at all, but proceeds to reject all further uses of impredicative definitions"(Wang 1974, 196).
[132] See (Solomon Feferman 1964) for an overview.

proof theory all attest to the fruitfulness of the underlying ideas of RTT from a purely mathematical standpoint, which must be beyond the expectation of Russell himself.

2.4.5 Simple Type Theory as a Theory of Concepts and Intensional Paradoxes

As we have argued before in 3.2, not only is STT a different system from RTT from a mathematical point of view, but also the underlying principles for them are totally independent. Unlike the constructivistic view of logical concepts and classes and the VCP, of which RTT is a natural consequence, Russell based his STT on three entirely different reasons. The first is that "it [STT] has a certain consonance with common sense which makes it inherently credible" (Whitehead and Russell 1927, 37). Just like we form clubs out of individuals, and form associations from clubs, logical entities form a similar hierarchy where types do not overlap and each of them is formed from entities just one type below. Different types of variables with disjoint ranges in mathematical practice also represent similar ideas. Secondly, propositional functions have a certain kind of ambiguity which can be eliminated only when combined with an appropriate argument. This idea is very close to Frege's doctrine of the "unsaturatedness" of concepts, a feature which had induced Frege to establish a theory of concepts very much like STT.[133] This idea is also reflected in Russell's view that a propositional function presuppose the totality of all its values, thus all its

[133] Thus Frege's system does avoid the intensional concept paradox, but cannot avoid the extensional class paradox because he operates with classes without any restriction. This point was already noticed by Frege himself in his famous letter to Russell, see (Frege 1902).

arguments. According to his conception, propositional functions with different arguments, i.e., different ambiguities, cannot replace each other, which is the basis of STT. Thirdly, a related but different reason is the idea of a "range of significance". Russell put it in the following way when he first proposed it:

> Every propositional function φ(x)—so it is contended—has, in addition to its range of truth, a range of significance, i.e., a range within which x must lie if φ(x)is to be a proposition at all, whether true or false. This is the first point in the theory of types; the second point is that ranges of significance form types, i.e., if x belongs to the range of significance of φ(x), then there is a class of objects, the type of x, all of which must also belong to the range of significance of φ(x), however φ may be varied. (Russell 1903, 523)

How convincing are Russell's reasons and how credible is STT as a theory of classes and concepts, apart from its power to solve all the known paradoxes and establish classical mathematics? To start with, STT does share some of our intuitions in ordinary talking, but it does in such a way that seems to be much too restrictive, as it forbids the formation of mixed type, for example. We can meaningfully talk about the classes of two persons like Gödel and Einstein and at the same time the classes of two universities or two countries that they all share some property (being a small set of two, for instance). This is related to Russell's second point that the same objection can be raised about the type distinctions for what we expect to be a concept (such as two)

would seem to be something realized in all the types[134] and not something limited in one particular type. According to this realistic understanding, we should treat concept as existing independently itself, just like the argument, rather than "fragments" of propositions, which don't have a meaning of their own, but only get a derivative one in forming propositions when combined with suitable arguments. In this aspect, Gödel would strongly oppose either Frege or Russell:

> 9.1.26 A concept is a whole—a conceptual whole—composed out of primitive concepts such as negation, existence, conjunction, universality, object (the concept of) concept, whole, meaning, and so on. We have no clear idea of the totality of all concepts. A concept is a whole in a stronger sense than set; it is a more organic whole, as a human body is an organic whole of its parts. (Wang 1996, 295)

Even if we take a constructivistic view about concepts as ambiguous or unsaturated entities, and assume at the same time that VCP in its second form (no totalities can contain entities involving this totality), STT doesn't follow from it in a strict sense in that mixed types don't contradict this form of VCP either, ZF then seems to be a natural extension of STT as a theory of classes. As for Russell's third reason, it's to be acknowledged that it does have an intuitive appeal when we say for example that "wisdom is square" is not false but meaningless, but the additional

[134] It's hard to insist, for example that the set of Gödel and Einstein falls under two, but that the pairing set of the unit set of Gödel and the unit set of Einstein doesn't fall under two, or falls under a different two.

condition that all meaningful arguments for any concept form a type seems very suspicious, since it is definitely conceivable that some concepts can apply to objects in different levels of the type hierarchy. Another difficulty of this principle is that, as Gödel put it, "its very assumption makes its formulation as a meaningful proposition impossible" (Gödel 1944, 138). The reason for this is that an equivalent formulation for Russell's principle is "whenever an object x can replace another object y in one meaningful proposition, it can do so in every meaningful proposition", which is either meaningful but trivial when x and y are of the same or non-trivial but meaningless when x and y belongs to different types. Another consequence is that an object x is (or is not) of a given type cannot be expressed by a meaningful proposition. When we try to apply the definition of being of the same type to a particular case, say, that Russell is of the same type as Socrates but not as courage, then we find ourselves involved in a contradiction. From our claim it follows that Socrates is not of the same type as courage, and yet our claim itself is a propositional function meaningfully applied to both Socrates and courage. Thus as a theory of concepts STT seems far less plausible than as a theory for classes (after eliminating the unnecessary restrictions), especially when concepts are conceived as something objective, which is very plausible on its own account in that some of them are be legitimately applied to itself, violating the type principle. This brings a central difference between classes and concepts, or for Gödel, mathematics and logic and the distinction between different types of paradoxes. In his conversations with Hao Wang, he said:

8.6.3 Sets and concepts are introduced differently: their connections are only outward. For instance, which no set can belong to itself, some concepts can apply to themselves; the concept of concept, the concept of being applicable to only one thing (or one object), the concept of being distinct from the set of all finite mathematical sets, the concept of being a concept with an infinite range, and so on. It is erroneous to think that to each concept there corresponds a set. (Wang 1996, 274)

Mathematics is mainly concerned with extensions, i.e., sets. Being a quasi-object, or so to speak, the limiting case of ordinary objects, sets cannot belong to themselves and the iterative conception of set resolves the set-theoretical paradoxes in a correct way, as we noticed earlier. While self-application or self-reference may cause a problem in the semantic paradoxes like the Liar Paradox, it always occurs in a definite language and we can correctly resolve it by a distinction of levels of language and use the solution to show positively that a complete epistemological description of a language within itself is impossible. In contrast, intensional paradoxes have nothing to do with language because they only involve logical concepts, such as the concept "not applying to itself"[135]. It should be noted, however that the third reason Russell gave in support of STT, after abandoning the extra condition that the range of significance forms types, brings in a new idea for the solution of the paradoxes, especially in their

[135] If we allow the objective existence of concepts and self-application of all concepts at the same time, then define $\Phi(\varphi) \leftrightarrow \neg\varphi(\varphi)$, substitute Φ for φ in the definition, we get $\Phi(\Phi) \leftrightarrow \neg\Phi(\Phi)$, a contradiction, which is the intensional form of Russell's paradox.

intensional form. It consists in not blaming the axiom that every propositional function defines a concept or class for the paradox, but rather the assumption that every concept gives a meaningful proposition for any arbitrary object as argument and this suggests a possible way to solve the intensional paradox. In the words of Gödel, "it might even turn out that it is possible to assume every concept to be significant everywhere except for certain 'singular points' or 'limiting points', so that the paradoxes would appear as something analogous to dividing by zero" (Gödel 1944, 138). A logical system of concepts based on such a principle would "be most satisfactory in the following respect: our logical intuitions would then remain correct up to certain minor corrections, i.e., they could then be considered to give an essentially correct, only somewhat 'blurred', picture of the real state of affairs"(Ibid.). The similar idea was suggested already by Frege in his proposed solution to the paradox in his solution, which proved only too primitive in that it can avoid the known paradox, but only at the cost of bringing back another one.[136] Concerning the idea of ranges of significance for concepts, an obvious objection can be raised that for every concept either total or partial[137], we can associate a total new one: if it is total we keep it unchanged, if it is partial, then we define the new one which gives a false proposition whenever the original one was meaningless, otherwise unchanged. This objection, however, can be met by pointing out that it is far from a certainty that we can always meaningfully determine whether a concept is total or partial, or, whether a

[136] See (Quine 1955) for details.
[137] Here we are borrowing some of the terminologies from computability theory. A concept is total if it is meaningful for every possible argument, otherwise partial.

proposition is meaningful when applying a concept to an arbitrary argument. A similar situation happens in recursive function theory too. All the recursive functions are either total or partial, but there is no recursive way to decide which of them are total and which are partial, otherwise we could easily use the diagonal trick to obtain a new recursive function. What's more, we cannot even extend recursively all the functions into total ones by filling the gaps of those partial ones by stipulation, i.e., there are always partial recursive functions which cannot be extended into a total one in a recursive manner.[138] To show that the intensional paradoxes are different from semantic ones and are a real threat for logic and conceptual realism, Gödel mentioned a briefer intensional paradox which he called "Church's Paradox", since it is most easily set up in Church's system (Church 1932, 1933) of function application. It may better be called "Gödel's Paradox", according to Wang (Wang 1996, 279)

A function is said to be regular if it can be applied to every entity (which can be an object or a function/concept). Consider the following regular function of two arguments:

(1) d(F,x)=F(x) if F is regular

=0 otherwise

Introduce now another regular function:

(2) E(x)= o, if x is not equal to 0,

[138] For a simple and elegant proof of this somewhat surprising fact, see theorem 43.9 in (Smith 2013, 336).

= 1 if x=0

We see immediately:

(3) E(x)≠x, (function E can be seen as a detecting whether the argument is 0 or not, or plays a similar role of the logical function of negation)

Let H(x) be E(d(x,x)), which is regular, since both E and d are regular functions. By (1), we have:

(4) d(H,x) = H(x) = E(d(x,x)).

Substituting H for x, we get:

(5) d(H,H) = E(d(H,H)), contradicting 3.

This paradox is simple yet striking in that the above derivation uses only the logical relation of predication (applying a concept to an argument) and definition by cases without even the need of propositional logic. It shows, on the one hand, that we don't yet perceive the concept of "concept" as clearly as in the other cases like sets. On the other hand, the existence of the intensional paradoxes may be used to show that concept has a non-arbitrary existence. As Gödel put it:

> 8.5.20 The intensional paradoxes … prove that we are not free to introduce any concepts, because, by definition, if we were really completely free, they [the new concepts] would not lead to contradictions. It is perfectly all right to form concepts in the familiar manner: we have evidence that these are meaningful and correct ways of forming concepts. What is wrong

> is not the particular ways of formations, but the idea that we can form concepts arbitrarily by correct principles. These principles are unavoidable: no theory of concepts can avoid them. Every concept is precisely defined, exactly and uniquely everywhere: true, false, or meaningless…. We don't make concepts, they are there. Being subjective means that we can form them arbitrarily by correct principles of formation. (Wang 1996, 273)

That concepts do not totally yield to our constructions, even if constructions following correct principles, do suggest they have robust existence just like physical objects. But this characteristic of concept is most clearly seen in a situation where seemingly totally different ways to capture it turn out to be equivalent, just like perceiving a physical object from different directions and positions. Fortunately we do have such a paradigm case, i.e., the notion of mechanical computability. The precise characterization of it as partial recursive functions should give us confidence in seeking a similar theory in the theory of concepts. More importantly, it will bring out the importance of the method of conceptual analysis, and, by delimiting the extent of pure mechanism and formalism, the role of intuition, to the discussion of which we will turn in the next chapter.

3. Computability in the Thirties: Gödel, Church, Turing and Beyond

3.1 Introduction

In the history and development of logic in the twentieth century, the thirties occupy a pivotal role: new ideas and concepts are developed, fundamental theorems proved and paradigmatic methods of proof codified and universally applied. On the list we can easily put, *inter alia*, Gödel's famous completeness theorem for first order logic (1930), his even more famous two incompleteness theorems for formal system of arithmetic (1931), Tarski's definition of truth (1935), Skolem's nonstandard model of arithmetic (1934), Gentzen's consistency proof for first-order Peano arithmetic via transfinite induction (1936) and Gödel's consistency proof for the axiom of choice and continuum hypothesis with Zermelo-Frankel set theory (1938). However, in terms of the robustness of the concept itself, the philosophical subtleties and its practical consequences, the clarification and all the continual development of the concept of effective computability and its formal counterpart, general recursive functions,[139] stands out as the single most important and interesting logical achievement. Emil Post,

[139] There has been a debate as to what should count as the best terminology of this logical discipline, whether recursive function theory or computability theory. Due mainly to the efforts of Soare (cf. (Soare 1996, 1999), who argues forcefully that from either the historical or philosophical point of view "computability theory" is the more appropriate one, most writers start using computability more often now, as can be seen from the title of the most recent textbooks on this subject. As for the purpose of our discussion, the difference isn't as salient as the art of naming a subject, due to the well-known equivalent proofs between Turing computability and general recursiveness.

who was one of the founders of this subject, wrote in 1944 that "indeed, if general recursive function is the formal equivalent of effective calculability, its formulation may play a role in the history of combinatory mathematics second only to that of the formulation of the concept of natural number" (Post 1944, 336). Gödel too, reflecting on the great importance of this concept, wrote that "this importance [of the concept of Turing computability or general recursiveness] is largely due to the fact that with this concept one has for the first time succeeded in giving an absolute definition of an interesting epistemological notion, i.e., one not depending on the formalism chosen" (Gödel 1946, 150). By a kind of miracle, four totally different attempts by Turing, Church, Kleene and Post[140] to characterize the informal notion of an effective method (or mechanical procedure/computable functions) in the same year 1936 all turn out to be extensionally equivalent. The claim that all effectively computable or mechanically computable functions are exactly the general recursive function or Turing computable functions is called the "Church-Turing Thesis" (CTT thereafter)[141] and is almost universally accepted today. However, just exactly what CTT is, or what kind of epistemological status it should have, is not as clear as the technical part of this whole subject. An especially interesting and illuminating point of view is Gödel's role in the history of the development.

Despite the appreciation of the importance of a rigorous notion of computability

[140] See Turing 1936; Kleene 1936; Church 1936b; Post 1936. It should be noted that Kleene 1936 is only a generalization and development of the ideas already present in Gödel 1934, and Post's work is independently of Turing, though he already knew Church's work when Post published his article.

[141] Church and Turing's paper and proposal both appeared in 1936. But the term "Church's Thesis" was used firstly by Kleene in (Kleene 1943), and "Turing's Thesis" in (Kleene 1952). The full term "Church-Turing Thesis" seems to have been first introduced by Kleene in (Kleene 1967, 232), with "a small flourish of bias in favor of Church" (Copeland 2002): "So Turing's and Church's theses are equivalent. We shall usually refer to them both as *Church's Thesis*, or in connection with one of its … versions which deal with 'Turing machines' as the *Church-Turing thesis*".

and his active role in the emergence of this notion, Gödel's part in its development is, at best "dichotomous", or as Dawson suggested, "at once seminal and indirect" (Dawson 2006, 133). On the one hand, although both Dedekind and Skolem had earlier studied and proposed the idea of defining arithmetical functions by recursion, it is Gödel who in the process of his incompleteness proof back in 1931 first systematically developed the theory of primitive recursive functions and predicates.[142] Then in 1934, in a series of lectures on undecidable propositions of formal mathematical system at the Institute for Advanced Study Gödel defined the notion of general recursive function[143], which served as the rigorous mathematical notion in Church's first published formulation of "Church's Thesis" in 1936, and which, together with another famous trick of the arithmetization of syntax also developed by Gödel in 1931, also formed the starting point for Kleene's further generalization and important theorems in recursive function theory. On the other hand, however, he never developed the theory in a systematic enough way for any theorems to bear his own name like other branches of logic.[144] And more importantly, although the technical machinery was already well in his hands, Gödel still refused or at least was reluctant to launch a "thesis" that would later bear Church's (and Turing) name, and more surprisingly, a thesis which he would eventually accept, due to Turing.[145] Thus, the

[142] One central step in the proof was to show that the metamathematical relation "x is a proof of y" is primitive recursive and thus representable in the formal system itself. Gödel simply called them "rekursiv". The term "primitive recursive" is, of course, a later invention.
[143] Gödel's definition was based on a suggestion of Jacques Herbrand in their correspondence; see (Sieg 2005) for a detailed analysis.
[144] One could well say the same about Gödel's role in the origin of computer science. He stated a "speed-up" theorem about the length of proof in 1936 and raised a question closely related to the P=NP problem in a letter to von Neumann in 1956, both of which are of central importance in computational complexity theory today. Yet, unlike Turing or von Neumann he had no real contribution in the emergence of digital computers.
[145] An extremely interesting comparison could be made here as to Gödel's "near miss" with CTT and Skolem's case

natural question for us is, why, from our hindsight, Gödel is only a bystander in the development of recursion theory, when he could well have done much better? Or, using Martin Davis's words, "Why Gödel didn't have Church's Thesis?" (Davis 1982) A related yet more concrete question is why Gödel finally accepted CTT, or on what grounds does he favor Turing's version, while rebutting Church's version, despite their formal and extensional equivalence? An even more puzzling problem is that, although on numerous occasions Gödel praised Turing's analysis for giving a "precise and unquestionably adequate definition" of formal system or mechanical computability, in a note published in 1972 Gödel claimed to have found a "philosophical error" in Turing's argument that mental procedure cannot go beyond mechanical ones. Our problem is then to reconcile these two seemingly inconsistent statements, i.e., try to understand how Gödel could enjoy the generality conferred on his incompleteness results by Turing's work, despite the error of its ways.

To gain a full understanding and possible solution of these three problems we must go back to their broad intellectual and foundational context. In section two I will outline the informal notion of an algorithm/mechanical computable function and why the search for the precise mathematical definition was needed in the bigger context of foundational researches in the early twentieth century. Section three will discuss the history of CTT and try to provide an answer to our first problem. Section four will deal

---

with completeness theorem of first order logic. Skolem failed to Gödel's remark that it was Skolem's philosophical "prejudice or blindness" towards non-finitary reasoning that caused him to fail to prove the theorem makes this comparison even more striking. Is it possible that he commits the same fallacy here? Dawson seems to support this explanation: "Until Turing's work Gödel resisted accepting Church's thesis and we may well wonder whether Gödel's focus on specific formalisms did not tend to blind him to the larger question of algorithmic undecidability. Gödel repeatedly stressed the importance of his philosophical outlook to the success of his mathematical endeavors; perhaps it may also have been responsible for an occasional oversight." (Dawson 1984) However, from our account later, this simplified view of Gödel is not tenable.

with the second problem, i.e., in what sense, if at all, is Turing's analysis more convincing than Church's? In section five I will turn to Gödel's criticism of Turing and see whether and how this could be consistent with his praise of Turing at the same time.

3.2 The Search for a Mathematical Definition of Effectively Computability

In this section I will discuss first the informal notion of an algorithm as it appears to be used in mathematical and logical contexts, and then give three different lines of motivation to explain why a search for a mathematical definition of this concept, which has been used in more than 2000 years in an informal way, suddenly becomes urgent and indicate the different working contexts of Turing, Gödel, and Church.

3.2.1 The Informal Notion of Algorithm

Mathematicians, from very early age on, or at least since Euclid, have experienced and discovered many algorithms such as the classical arithmetical method of finding the greatest common divisor of two integers or the algebraic method of finding

solutions for quadratic equations. Even without a precise definition of what an algorithm is, they have never failed in recognizing one when an effective method really occurs. The search for algorithms is always connected with the aim of finding a universal method for an infinite class of problems of similar character, i.e., a special case for taming the infinity with the finite. An algorithm or computation procedure in the intuitive sense must be fully and finitely described before any particular question is selected out of a potential infinite number to which it is applied. When the question has been selected, all steps must then be predetermined and performable without any exercise of ingenuity or mathematical invention by the person doing the computing. Abstracting from these commonalities we could say that a method, or procedure M for achieving a desired outcome is called 'effective' or 'mechanical'[146] in the informal sense just in case[147]

1. M is set out in terms of a finite number of exact instructions (each instruction being expressed by means of a finite number of symbols);
2. M will, if carried out without error, produce the desired result in a finite number of steps;
3. M can (in practice or in principle) be carried out by a human being unaided by any machinery save paper and pencil;
4. M demands no insight or ingenuity on the part of the human being

[146] As we can see later Gödel had reservations equating "mechanical" with "effective" as is normally assumed, via his speculation of an "effectively but non-mechanical procedure". However, apart from the particular discussion relating to Gödel, we will use these two words interchangeably in the text.
[147] We are following (Copeland 2002) in giving these descriptions. There are lots of other characterizations by similar ways, see for example (Kleene 1952, 1967)

carrying it out.

Although these four conditions as a whole still don't, possibly, narrow the informal notion down to a precise mathematical notion, we have gone a much further way in making it sharper.[148] First, in requiring the set of instructions and steps for carrying out the method to be finite without giving any fixed bound beforehand it can be seen clearly that here we are dealing with an idealized notion of effective method, ignoring any feasibility conditions about the length, the complexity or the memory or time required of the execution of the method. Secondly, we are explicitly discussing the extent and limit of human computability, i.e., what can be effectively or mechanically done by a human calculator ignoring time and energy or other irrelevant factors, thus setting aside the question about what could (not) be done by an arbitrary machine thought of as a physical system. That is to say, we assume that CTT is a thesis about human beings rather than with machines in general.[149] The last and also maybe the most important and salient feature of an effect or mechanical feature is that it allows the human computer to proceed in a routine or algorithmic fashion, without the extra

[148] The much debated, or as we could say, the dominant problem concerning CTT in literature about the (un)provability of CTT hinges on the crucial problem of how to correctly understand this informal part of the equivalence. For someone who thinks that the informal notion of an effective method is always a concept of "open texture", a "proof" of mathematical rigor could indeed never be found. However we won't go directly into this problem about the provability of CTT in our discussion below, though we will touch aspects of it via Gödel's concept of it. For further arguments about it see (Shapiro 2013) and Smith's "squeezing argument" against the idea of an indefinite uncertainty of this open texture concept in (Smith 2013§45) and (Mendelson 1990).

[149] Again the problem of whether CTT is only a relatively modest (but still quite ambitious!) claim about human calculations or a much wider claim about humans and machines in general is a much discussed issue, especially since Gandy's illuminating paper (Gandy 1980) to separate a human version of Turing's thesis and the related yet distinct machine version of the thesis. (Hodges 2006) and (Copeland 2006) are the representatives of each side and have both argued forcefully for their own views. I tend to support Copeland in this issue, both from a historical and philosophical perspective. The problem Church and Turing have in mind, at least in 1936, is about what can be humanly effectively done and their argument (especially Turing's) may or may not apply equally to the case of machine computability. But I also agree with Shagrir that when CTT is used in the context of computer science or physical computation or neuroscience rather than a philosophical discussion today, the machine version is more often implicitly assumed. See (Shagrir 2002).

need of any insight or ingenuity, or equivalently, without having recourse to the meaning of the symbols under operation apart from those syntactical features like shape or order.[150] We should say a few more about condition 2 that an effective method will always terminate in a finite number of time and steps to produce an result, it is actually too restrictive. It may be argued that a procedure which sometimes fails to give a result is a mathematically uninteresting and therefore artificial procedure lacking any real value. However, we could also argue that either mathematically or intuitively that it's not in the idea of an effective procedure that it always terminates. Intuitively, as long as a method allows us to proceed in each step mechanically, it deserves the name of an effective procedure. Mathematically, the totality of computable functions become a stable unity, i.e., not susceptible of diagonal argument as it happens in the primitive recursive functions, only because the introduction of partial recursive functions, functions which are undefined for certain of their arguments.[151] This is pointed out clearly by Gödel, as reported by Hao Wang:

> Gödel points out that the precise notion of mechanical procedures is brought out clearly by Turing machines producing partial rather than general recursive function. In other words, the intuitive notion does not require that a mechanical procedure should always terminate or succeed. A sometimes unsuccessful procedure, if sharply defined, still is a procedure, i.e., a well

[150] We will see later that for Gödel it's precisely because the possibility of grasping the abstract meaning of symbols would still count as an "effective" method that he distinguishes the "effective" from the "mechanical".
[151] Or, if we insist that effective functions are those total recursive functions, then unlike the more natural set of partial recursive functions, we cannot even list them in any effective way, otherwise the diagonal method easily applies.

> determined manner of proceeding. Hence we have an excellent example here of a concept which did not appear sharp to us but has become so as a result of a careful reflection. The resulting definition of the concept mechanical is both correct and unique. Unlike the more complex concept of always-terminating mechanical procedures, the unqualified concept, seen clearly now, has the same meaning for the intuitionists as for the classicist. Moreover it is absolutely impossible that anybody who understands the question and knows Turing's definition should decide for a different concept. (Wang 1974, 84)

3.2.2 Effective Computability and Entscheidungsproblem

The most direct impetus for a precise definition of an algorithm comes from a problem in logic: the decision problem for first order logic, and it is this problem that leads Turing to his own solution. In what is arguably the first modern presentation of first order logic, Hilbert formulated the decision problem for predicate logic as follow: "The Entscheidungsproblem is solved if one knows a procedure that permits the decision concerning the validity, respectively, satisfiability of a given logical expression by a finite number of operations. …We want to make it clear that for the solution of the decision problem a process needs to be given by which validity can, in principle, be determined (even if the laboriousness of the process would make using it impractical)" (Hilbert and Ackermann 1950, § 12).

The original formulation is a model-theoretic one, the more familiar, and equivalent, via the completeness theorem, proof-theoretic one as formulated by Church is: "By the Entscheidungsproblem of a system of symbolic logic is here understood the problem of finding an effective method by which, given any expression Q in the notation of the system, it can be determined whether or not Q is provable in the system"[152] (Church 1936a, 112–13). It is to be noted that Hilbert characterized the problem as the "fundamental problem" of mathematical logic, because it seemed plausible for him that if the solution of this problem can be found, it should also be possible to solve all the other important mathematical questions in a purely mechanical manner, at least in principle.[153] Hence, given some mathematically unsolvable problem at all, then the *Entscheidungsproblem* itself should be unsolvable. Although some of the subclasses of first order sentences have been shown to be decidable, unlike the completeness problems posed in the same book, a negative solution for the decision problem is expected, as can been seen from the views expressed already in 1927 by von Neumann as follows:

> … it appears that there is no way of finding the general criterion for deciding whether or not a well-formed formula a is provable. (We cannot at the moment establish this. Indeed, we have no clue as to how such a proof of

[152] Note that the two senses of Entscheidungsproblem are equivalent in view of the completeness of first order logic.
[153] To be more precise: if some important mathematical theories can be finitely axiomatized in a formal system, then whether any particular mathematical theorem is true in that theory can be transformed into a question of whether any particular implication sentence is logically valid via the deduction theorem, and thus become decidable too.

> undecidability would go.)… the undecidability is even a *conditio sine qua* non for the contemporary practice of mathematics, using as it does heuristic methods, to make any sense. The very day on which the undecidability does not obtain any more, mathematics as we now understand it would cease to exist; it would be replaced by an absolutely mechanical prescription by means of which anyone could decide the provability or unprovability of any given sentence. … Thus we have to take the position: it is generally undecidable, whether a given well-formed formula is provable or not. (von Neumann 1927, 11–12, translation from Sieg 1994)

While expecting confidently a negative solution, even von Neumann has "no clue as to how such a proof of undecidability would go". It only shows that the underlying conceptual difficulty here is of a different order compared with traditional problems. Just how could a proof of undecidability be given? The unsolvability results of other mathematical problems in history had always been established relative to a determinate class of admissible operations, e.g., the impossibility of doubling the cube relative to ruler and compass constructions or the impossibility of squaring the circle. A negative solution to the decision problem obviously required a mathematically precise characterization of all "effective procedures". Or to quote Kleene:

> The intuitive notion of a computation procedure, which is real enough to separate many cases where we know we do have a computation procedure

> before us from many others where we know we don't have one before us, is vague when we try to extract from it a picture of the totality of all possible computable functions. And we must have such a picture, in exact terms, before we can hope to prove that there is no computation procedure at all for a certain function, or briefly to prove that a certain function is uncomputable. Something more is needed for this. (Kleene 1967, 32)

Like the classical unsolvability proofs, these proofs are of unsolvability by means of given instruments. But where the analogy fails and what is really new in the decision problem is that in the present case we are seeking the impossibility of something by means of all the instruments at human beings' disposal. The impossibility proofs for angle trisection and circle squaring are for constructions using specific instruments such as straightedge and compass and this alone makes the result relative; using more powerful instruments, there is no difficulty with either of these geometrical construction problems. Matters are different with the decision problem for first order logic; here what we will show is that there is no solution using any methods whatever available to human beings, a much more difficult problem and whose solution has a much more significant implication.

3.2.3 Effective Computability and the Scope of the Incompleteness Theorems;

With hindsight it is difficult to doubt the universal applicability and the significance of Gödel's incompleteness theorem. However, in the early thirties immediately after Gödel announced his incompleteness theorem, the situation was totally different. Eminent logicians, such as Alonzo Church, tended to avoid the unpleasant consequence of incompleteness theorems by trying to restricting the range of its applicability, as can be seen clearly from a letter Church wrote to John Dawson in 1983 responding to the latter's inquiry, whether Church had been "among those who thought that the Gödel incompleteness theorem might be found to depend on peculiarities of type theory". Church's answer goes like this:

> "… yes I was among those who thought that the Gödel incompleteness theorem might be found to depend on peculiarities of type theory (or, as I might later have added, of set theory) in a way that would show his results to have less universal significance than he was claiming for them. There was a historical reason for this, and that is that even before the Gödel results were published I was working on a project for a radically different formulation of logic which would (as I saw it at the time) escape some of the unfortunate restrictiveness of type theory. In a way I was seeking to do the very thing that Gödel proved impossible, and of course it's unfortunate that I was slow

to recognize that the failure of Gödel's first proof to apply quite exactly to the sort of formulation of logic I had in mind was of no great significance"[154] (Sieg 1997, 177).

Gödel's results about undecidable propositions were first published in 1931 under the title "On formally undecidable propositions of *Principia Mathematica* and related systems I" (Gödel 1931). As the title indicates, the results apply to "*Principia Mathematica* and related systems," and, more precisely, to the formal system P, which is "essentially the system obtained when the logic of PM is superposed upon the Peano axioms" and its extensions, which are the "ω-consistent systems that result from P when [primitive] recursively definable classes of axioms are added" (Gödel 1931, 185). In his 1934 Princeton lectures he already realized this problem of generality, and gave the title of his series of lectures "On Undecidable Propositions of Formal Mathematical System", trying to incorporate as many formal mathematical systems as possible. He attempts to characterize "formal mathematical system" by requiring that the rules of inference and definitions of meaningful [i.e., syntactically well-formed] formulas and axioms be "constructive" by emphasizing the finite nature of formal systems: "We require that the rules of inference, and the definitions of meaningful formulas and axioms, be constructive; that is, for each rule of inference there shall be a finite procedure for determining whether a given formula B is an immediate

[154] "No great significance" that Church referred to in his letter was probably the fact that his original logical system was inconsistent, discovered by his pupil Kleene and Rosser, thus escaping Gödelian incompleteness, but in the least interesting way. For more information about the reception of Gödel's incompleteness theorem in their contemporary context, see (Dawson 1984).

consequence (by that rule) of given formulas $A_1, \ldots A_n$, and there shall be a finite procedure for determining whether a given formula A is a meaningful formula or an axiom" (Gödel 1934, 346). In section 6 where he explicitly discusses the "conditions that a formal system must satisfy in order that the foregoing arguments [the proof outline of the first incompleteness theorem] apply", Gödel lists as the first condition that "the class of axioms and the relation of immediately consequence shall be recursive"[155] and he adds that "this is a precise condition which in practice suffices as a substitute for the unprecise requirement that the class of axioms and the relation of immediate consequence be constructive"[156]. The "mechanical" aspect for a formal mathematical system was emphasized by Gödel in his 1933 lecture on the "present situation in the foundation of mathematics". After an explanation for what a formalization of mathematics means, i.e., setting up a perfectly precise language to possibly express any mathematical proposition by a formula in the language and laying down certain of the formulae as axioms and certain rules of inference which allow one to pass from the axioms to new formulas and thus to deduce more and more propositions, he pointed out that "the outstanding feature of the rules of inference being that they are purely formal, i.e., refer only to the outward structure of the formulas, not to their meaning, so that they could be applied by someone who knows nothing about mathematics, or by a machine [This has the consequence that there can never be any doubt as to what cases the rules of inference apply to, and thus the

[155] "Recursive" for Gödel here refers to primitive recursive.

[156] As we know now, by Craig's theorem (see (Putnam 1965) for an wonderful exposition), that any recursively enumerable axiomatized theory can be recursively axiomatized, and even in a primitive recursive way. So Gödel was right after all.

highest possible degree of exactness is obtained.]" (Gödel 1933c, 45). As we can see, the essential difficulty in an adequate definition of a formal system is the notion of a finite and mechanical procedure, since with this notion in hand we can decide what is a well formed formula and which formulas are the axioms and also since a rule of inference is nothing else but a mechanical procedure which allows one to determine of any given finite class of expressions whether anything can be inferred from them by means of the rule of inference under consideration, and if so to write down the conclusion. Furthermore, the connection between a mechanical procedure in the case of formal systems with computable functions leap to the eye, since expressions or finite classes of expressions can be coded by natural numbers, and therefore a procedure in the sense of an inference rule is nothing else but a function $f(x_{1,\ldots,}x_n)$ whose arguments as well as its values are natural numbers and which is such that for any given system of natural numbers the value can actually be calculated. This includes also the case of deciding the axioms where the answer Yes or No is to be obtained by a procedure, because Yes or No can be replaced by the integers 0 and 1. So the concept of mechanical procedure to be specified is the same concept of a calculable function of integers. In an unpublished paper, probably from 1938, Gödel writes:

> When I first published my paper about undecidable propositions the result could not be pronounced in this generality, because for the notions of mechanical procedure and of formal system no mathematically satisfactory definition had been given at that time. The gap has since been filled by

Herbrand, Church and Turing. (Gödel 193?, 168)[157]

More explicitly, in his Postscriptum to his 1931 paper, and in a more detailed form in his famous Postscriptum to his 1934 Princeton lectures, Gödel says:

> In consequence of later advances, in particular of the fact that, due to A. M. Turing's work, a precise and unquestionably adequate definition of the general concept of formal system can now be given, the existence of undecidable arithmetical proposition and the non-demonstrability of the consistency of a system in the same system can now be proved rigorously for **every** consistent formal system containing a certain amount of finitary number theory.
>
> Turing's work gives an analysis of the concept of "mechanical procedures" (alias "algorithm" or "computation procedure" or "finite combinatorial procedure"). This concept is shown to be equivalent with that of a "Turing machine." A formal system can simply be defined to be any mechanical procedure for producing formulas, called provable formulas. For any formal system in this sense there exists one in the sense … above that has the same provable formulas (and likewise vice versa), provided the term "finite procedure" … is understood to mean "mechanical procedure". This

[157] Martin Davis dates the article to 1938, see his introduction to Gödel 193?. The mention of Herbrand and Church along with Turing should not mislead us into thinking that for Gödel there's no difference in these definitions and this way of talking counts as a counter evidence for his favor on Turing. On the contrary, just a few pages later, he says explicitly that "That this [general recursive functions] really is the correct definition of mechanical computability was established beyond any doubt by Turing".

> meaning, however, is required by the concept of formal system, whose essence it is that reasoning is completely replaced by mechanical operations on formulas. (Note that the question of whether there exist finite non-mechanical procedures not equivalent with any algorithm, has nothing whatsoever to do with the adequacy of the definition of "formal system" and of "mechanical procedure".) (Gödel 1934, 369–70)

This analysis should have made it clear that, unlike Turing who was working in the context of finding a mechanical procedure for solving the decision problem for first order logic, Gödel's concern and orientation towards the definition of a finite, mechanical procedure was much different. This will affect our understanding later of how to read correctly Gödel's seemingly incompatible assessment of Turing's contribution. Already in the quotation above, we can get some hints of Gödel's ambivalence. In a footnote he refers to his 1958 article of extension of methods from "a finitary point of view", where he posits procedures that can be construed as fulfilling the finiteness requirement, in that they are constructive, yet are non-mechanical in that they "involve the use of abstract terms on the basis of their meaning". As we will show later, these procedures play an important role in Gödel's critique of Turing.

Effective computability and the Unsolvability of Mathematical Problems

The third and last context in which a precise definition of an effective or mechanical procedure is indispensable concerns certain mathematical problems. Most famous among them is the so-called "Hilbert's Tenth Problem", a problem that Hilbert lists in tenth place among 23 unsolved mathematical problems in his lecture at the Paris conference of the international Congress of Mathematics in 1900: Given a Diophantine equation with any number of unknown quantities and with rational integral numerical coefficients: to devise a process according to which it can be determined by a finite number of operations whether the equation is solvable in rational integers. (Hilbert 1900a)

Obviously in order to show that it is unsolvable we need to know in a mathematically precise sense what are the totality of all mechanical or effective processes.[158] Other mathematical problems which have a similar character are mentioned by Church and are exactly the context in which Church proposed his definition of effective calculability and his "thesis". In the words of Church, "There is a class of problems of elementary number theory which can be stated in the form that it is required to find an effectively calculable function F of positive integers, such that

[158] This problem was finally solved in 1970 (during the cold war) by joint efforts of a Russian mathematician Yuri Matiyasevich and three other American mathematicians Julia Robinson, Martin Davis and Hilary Putnam. The negative solution to Hilbert's tenth problems is thus known as the "MRDP Theorem". It's solution relies on CTT and the key step is to show that every recursively enumerable set is Diophantine and vice versa, together with the well-known result that there exists recursively enumerable but not recursive sets. For details see (Davis 1973; Matiyasevich 1993).

f(x)=2[159] is a necessary and sufficient condition for the truth of a certain proposition of elementary number theory involving x as free variables" (Church 1936b, 89). For example, the famous problem of Fermat's conjecture[160] to find a means of determining of any given positive integer n whether or not there exists positive integers x, y, z such that $x^n + y^n = z^n$ could be seen as such a problem, since we can interpret the original problems to be the requirement to find an effectively calculable function f, such that f(n)=2 if and only if there exist positive integers x, y, z such that $x^n + y^n = z^n$. Clearly the condition that f is effectively calculable is essential; otherwise it becomes a trivial problem. Other problems mentioned by Church include the topological problem "to find a complete set of effectively calculable invariants of closed three-dimensional simplicial manifolds under homeomorphisms" (ibid.). Important algebraic problems like the decision problem of the word problem for semigroups also belong to this type of problems. All these examples just show that apart from purely logical considerations, natural and interesting mathematical problems also call for a need for a mathematical definition of "effective calculable".

3.3 A Brief History of CTT[161]

[159] As noted by Church himself, the choice of 2 is not essential, it's just a "characteristic" number representing the truth/falsity of the number-theoretic proposition.
[160] Still a conjecture at the time of Church's writing, but a theorem since 1995.
[161] Lots of comprehensive and interesting articles have been written about the early history of CTT and other related matters, see especially (Gandy 1988; Sieg 1994; Davis 1982) For an account of the development of the concept of computability within a broader historical context, see (Sieg 1994, 2009). My account draws on them, but differs in the sense that I will consider the story more from Gödel's point of view.

In this section I will first outline the development of the theory of computable functions in Princeton, focusing on Church, Kleene and Gödel; then describe briefly the almost simultaneous work of Turing in Cambridge. After this we will consider Gödel's role in this whole development and his change of attitude after Turing's work and try to explain why Gödel didn't have CTT.

3.3.1 The Princeton Side

In the early1930s, as an effort to eliminate the cumbersome feature of type distinction in Russellian type theory and to avoid the "unpleasant" (Church 1934, 360) consequence of Gödel's incompleteness theorem that symbolic logic is essentially incomplete, Church had been proposing a new foundation for logic and mathematics in his 1932 and 1933 paper "A set of postulates for the foundation of logic" (Church 1932, 1933). Although his system was proved to be inconsistent later by his own students Kleene and Rosser (Kleene and Rosser 1935), an extraordinarily fruitful ingredient remained in this system: the so-called λ-calculus and, correspondingly the λ-definable functions.[162] Church's student, Kleene, showed by 1933 that the theory of positive integer and a large class of number theoretic functions were λ-definable.[163] What's more, apart from lots of familiar number-theoretic functions, effective operations like composition of functions, definition by primitive recursion, and most

[162] For a wonderful introduction to the basic ideas of λ-definable functions, see (Kleene 1981a).
[163] The same for Peano's axioms for natural numbers. The key step to this is the λ-definition of predecessor function, which was solved by Kleene, obviously "while in a dentist's office" (Kleene 1981a, 56).

interestingly the least-number operator, generally known as the μ-operation, are also λ-definable. Based on the strength of this evidence, Church had the idea that the notion of "effectively calculable" can be identified with "λ-definable", and proposed it to Gödel during his visit to the Institute for Advanced Study in Princeton in the fall of 1933. In the words of Kleene, "The concept of λ-definability existed full-fledged by the fall of 1933 and was circulating among the logicians at Princeton. Church had been speculating, and finally definitely proposed, that the λ-definable functions are all the effectively calculable functions—what he published in 1936, and which I in 1952 Chapter XII called 'Church's thesis'" (Kleene 1981a, 59). Adding to the evidence of the plausibility of Church's proposal is the amazing fact that, unlike the case of primitive recursive functions, the usual diagonalization procedure doesn't lead out of the class of the λ-definable functions, a fact which persuaded Kleene immediately: "When Church proposed this thesis, I sat down to disprove it by diagonalizing out the class of the λ-definable functions. But, quickly realizing that the diagonalization cannot be done effectively, I became overnight a supporter of the thesis" (Ibid.).

Church's encounter with Gödel, where he proposed the idea of identifying λ-definable with effectively calculable functions and Gödel's reaction was vividly expressed in a letter from Church to Kleene, dated November 29, 1935:

> In regard to Gödel and the notions of recursiveness and effective calculability, the history is the following. In discussion with him the notion

> of lambda-definability, it developed that there was no good definition of effective calculability. My proposal that lambda-definability be taken as a definition of it **he regarded as thoroughly unsatisfactory**. I replied that if he would propose any definition of effective calculability which seemed even partially satisfactory I would undertake to prove that it was included in lambda-definability. **His only idea at the time was that it might be possible, in terms of effective calculability as an undefined notion, to state a set of axioms which would embody the generally accepted properties of this notion, and to do something on that basis.** Evidently it occurred to him later that Herbrand's definition of recursiveness, which has no regard to effective calculability, could be modified in the direction of effective calculability, and he made this proposal in his lectures. At that time he did specially raise the question of the connection between recursiveness in this new sense and effective calculability, but he said he did not think that the two ideas could be satisfactorily identified except heuristically. [My own emphasis] (Davis 1982, 9)

"Herbrand's definition of recursiveness" that Church mentioned in this letter was an attempt by Gödel in his 1934 lectures to capture "what one would mean by 'every recursive function'" (Gödel 1934, 368). In section 2 of his lecture discussing "Recursive functions and relations" Gödel pointed out that " [Primitive] recursive functions have the important property that, for each given set of values of the

argument, the value of the function can be computed by a finite procedure" (Ibid. 348). There is also a suggestive footnote:

> The converse seems to be true, if besides [primitive] recursions … recursions of other forms (e.g., with respect to two variables simultaneously) are admitted. This cannot be proved, since the notion of finite computation is not defined, but it serves as a heuristic principle. (Ibid. footnote 3)

The wording of this comment could easily give us the impression that this footnote together with the proposed definition of general recursive functions in the same lecture amounted to a precise characterization of the effectively calculable functions and hence to an anticipation of a statement of Church's thesis. However, in a reply to Davis's query concerning this point Gödel pointed out clearly that:

> …it is not true that footnote 3 is a statement of Church's Thesis. The conjecture stated there only refers to the equivalence of "finite (computation) procedure" and "recursive procedure." However, I was, at the time of these lectures, not at all convinced that my concept of recursion comprises all possible recursions; and in fact the equivalence between my definition and Kleene's is not quite trivial.[164] (Davis 1982, 8)

[164] That is, the equivalence of general recursive functions (Gödel) with μ-recursive functions. (Kleene 1936) Due to the normal form theorem in Kleene 1936, stating that "each general recursive function is obtainable using only primitive recursions (with explicit definitions) and the least-number operator (used just once)" (Kleene 1981, 60) and the equivalence results mentioned above, Gödel's thesis (GT) should be very plausible. Martin Davis thus

In the light of Gödel's answer we can at most say that Gödel had in mind a particular version of Church's thesis:

Gödel's Thesis (GT): *Every mechanically calculable function can be defined using recursions of the most general kind.*[165] (Davis 1982, 6)

Gödel's attempt to define "recursions of the most general kind" leads to his definition of general recursive functions in section 9 of his 1934 lecture. The standard and the easiest way of defining a function by recursion is the schema of primitive recursion:

(a) $f(0,y) = h(y)$,

$f(n+1, y) = g(n, f(n, y), y)$

where g and h are supposed to be previously defined calculable functions, then f will likewise be calculable. Another common schema is by functional composition:

(b) $f(n) = h(g(n))$,

---

writes: "This theorem [normal form theorem] … must have gone a considerable distance towards convincing Gödel that his concept of recursion indeed 'comprises all possible recursions'." However, Gödel never mentioned Kleene 1936 or normal form theorem in his discussions about Church's thesis, and the speculation of "must have gone" at best will be "should have gone". See (Webb 1990, 294) for further remarks about the role normal form theorem played in convincing Gödel about the correctness of GT. For Kleene's argument that parital recursive functions include functions using all possible kinds of recursions, see (Kleene 1981b).

[165]See (Webb 1980, 186–203) for further discussion of this "thesis" and its root in ideas of Skolem and Hilbert.

f will be calculable if h and g are already calculable.

However, apart from the easily constructed diagonal function, Ackermann's function[166] already shows that the schema of primitive recursion (besides certain initial functions) is not strong enough to capture all intuitively effectively calculable functions. The suggestion from Herbrand (as reported by Gödel himself) is as follows: "If *f* denotes an unknown function, and $g_1, \ldots, g_k$ are known functions, and if the g's and the f are substituted into one another in the most general fashions and certain pairs of the resulting expressions are equated, then if the resulting set of functional equations has one and only one solution for f, f is a recursive function." (Gödel 1934, 368) The basic idea is to permit calculation of the value of a function for a particular argument by using equations connecting the value of the function with the value of other defined calculable functions and the other values of the same function in the most general conceivable manner.

Thus for example we might have

$f(x,0)= g_1(x)$

$f(0, y+1)= g_2(y)$

$f(1, y+1)=g_3(f(0, y), y)$

The two conditions Gödel added were that (1) the defining equations must be of a certain order and normal from and (2) that for each set of arguments x_n of the unknown

[166] See (Ackermann 1928) and a simplified form given in (Péter 1936).

function f there is exactly one value m such that $f(x_n)=m$ is a "derived equation", by certain elementary specifiable rules. These rules are basically substituting particular integers for the variables and to substitute equals for equals. More precisely, they are

(1a) Replacing all the variables of a given equation by numerals shall be a derived equation;

(1b) $f(k_{1,,,} k_n)=m$ shall be a derived equation if the k_i's are numerals and it is a true equality;

(2a) If for any of the known or auxiliary g_i, $g_i(k_{1,,,} k_n)=m$ is a derived equation, then the equality obtained by substituting m for $g(k_{1,,,} k_n)$ is a derived equation;

(2b) If for the unknown f and some particular numeral, $f(k_{1,,,} k_n)=m$ is a derived equation, then the expression obtained by substituting m for an occurrence of $f(k_{1,,,} k_n)$ on the right-hand side of a derived equation shall be a derived equation.

We take great pains to spell out the notion of "general recursive function", not only because this notion is an extremely natural one to generalize over primitive recursion to include all possible kinds of recursions and later turns out to be equivalent to Turing's machine computability, but also because Gödel's contribution in the discovery of this notion is unclear given that he says it's "suggested by Herbrand" and "essentially Herbrand" (Gödel 193?, 167).[167] On the one hand, apart from isolating this notion from the finitary epistemologically consideration with which Herbrand

[167] For a detailed analysis of Herbrand's suggestion and Gödel's contribution, see (Sieg 2005).

associated it, making it in the direction of a theory of computation, the importance of "normalizing" general recursive functional equations by laying down specific rules was that it allowed the arithmetization of the entire theory of general recursive functions. Only then, by treating the defining equations formally as sequences of symbols could Kleene later in 1936 expand the theory and get important and substantial results like the normal form theorem. On the other hand, the rules gave a deeper understanding of the nature of computation. In a letter to van Heijenoort on 23 April 1963 answering a question about the relation of Herbrand's equation and his definition of general recursive functions, Gödel wrote that "What he [Herbrand] failed to see (or to make clear) is that the computation, for all computable functions, proceeds by exactly the same rules. It is this fact that makes a precise definition of general recursiveness possible" (Gödel 2003b, 308). This uniform nature of computation occupies an important part in understanding it, as we will see later when Turing transformed the problem of what is a computable function or computation to the related yet more precise question of what are the possible processes which can be carried out in computing.

Despite the discouraging reaction from Gödel about his original proposal of identifying effective calculability with λ-definability, Church still goes on a year later to announce his "thesis" to the mathematical world. In a meeting of the American Mathematical Society in New York City on April 19, 1935, he gave a ten minute contributed talk. The abstract of that paper (received by the Society on March 22,

1935) is:

> Following a suggestion of Herbrand, but modifying it in an important respect, Gödel has proposed a definition of the term recursive function, in a very general sense. In this paper a definition of recursive function of positive integers which is essentially Gödel's is adopted. And it is maintained that the notion of an effectively calculable function of positive integers should be identified with that of recursive function, since other plausible definitions of effective calculability turn out to yield notions which are either equivalent to or weaker than recursiveness. There are many problems of elementary number theory in which it is required to find an effectively calculable function of positive integers satisfying certain conditions, as well as a large number of problems in other fields which are known to be reducible to problems in number theory of this type. A problem of this class is the problem to find a complete set of invariants of formulas under the operation of conversion. It is proved that this problem is unsolvable, in the sense that there is no complete set of effectively calculable invariants. (Church 1935)

The interesting point here is that Church changed the definitions from "λ-definable" to "recursive", his abbreviation for Herbrand-Gödel general recursive in the formal part of his thesis and λ-definability occurs only in the reference to "notions which are either equivalent to or weaker than recursiveness". This fact even leads

Martin Davis to conjecture that "in the early spring of 1935 Church was not yet certain that λ-definability and Herbrand-Gödel general recursiveness were equivalent". Even in his official 1936 paper "An unsolvable problem of elementary number theory", Church still used general recursiveness as a definition of effective calculable:

> We now define the notion, already discussed, of an effectively calculable function of positive integers by identifying it with the notion of a recursive function of positive integers (or of a λ-definable function of positive integers). This definition is thought to be justified by the considerations which follow, so far as positive justification can ever be obtained for the selection of a formal definition to correspond to an intuitive notion. (Church 1936b, 100)

At this time Church was definitely fully aware of the equivalence of λ-definability and recursiveness. In note 3, he mentions the different contributions of him of his collaborators: "The notion of λ-definability is due jointly to the present author and S. C. Kleene…the notion of recursiveness is due jointly to Jacques Herbrand and Kurt Gödel. And the proof of equivalence of the two notions is due chiefly to Kleene, but also partly to the present author and to J. B. Rosser. The proposal to identify these notions with the intuitive notion of effective calculability is first made in the present paper" (Ibid.). Footnotes 16 and 17 tell us that it's Kleene who proved the theorem that every recursive function of positive integers is λ-definable, Church and Kleene proved

independently that every λ-definable function of positive integers is recursive at about the same time. From this evidence we can reasonably say that in 1935 Church only knows that every λ-definable is recursive, thus he announced his earlier thesis in terms of recursiveness. Only later in 1935 due to a proof of Kleene of the other direction did he realize that λ-definable and recursive are co-extensive and adopted both formulations which are "equally natural" according to him in his thesis, and even took his co-extensiveness as evidence for the correctness of his own thesis: "The fact, however, that two such widely different and (in the opinion of the author) equally natural definitions of effective calculability turn out to be equivalent adds to the strength of the reasons adduced below for believing that they constitute as general a characterization of this notion as is consistent with the usual intuitive understanding" (Ibid. 90). As for Gödel's role in this whole development, he writes in footnote 18 that

> The question of the relationship between effective calculability and recursiveness (which it is here proposed to answer by identifying the two notions) was raised by Gödel in conversation with the author. The corresponding question of the relationship between effective calculability and λ-definability had previously been proposed by the author independently.

We can thus conclude that in 1936 in the search of a definition of effective calculable, on the one hand, for Gödel, he is "not at all convinced" that his own

general recursive functions include all possible recursion and the proposal of λ-definable is "thoroughly unsatisfactory", while on the other hand, for Church (and possibly Kleene), both definitions are "equally natural" and the inductive evidence that lots of mathematical functions can be defined in these two different formalism and the remarkable fact that these two "widely different" definitions turn out to be equivalent adds to the his conviction that either of them can be offered as the correct definition of the informal notion of effective calculability.

3.3.2 The British Side[168]

Unlike the situation in Princeton where the founders of λ-definable functions and general recursive functions could exchange ideas and discuss related problems (though they might be in huge disagreement with each other), the creator of the "Turing machine",[169] Alan Turing, had to work his own way into the problem. Earlier in 1935, still being a fellow of King's College Cambridge at the age of 23, Turing had attended a course on the Foundations of Mathematics given by the topologist M. H. Newman. Among other things, Newman explained Hilbert's problems concerning consistency, completeness, and decidability of various formal axiomatic systems, as well as Gödel's incompleteness results. Turing had already been interested in mathematical logic but

[168] See (Hodges 1983) for more details about Turing's life and work.
[169] Turing used a-machine (automatic) in his original 1936 paper. The famous name "Turing machine" first appeared in a review of Turing's paper by his then supervisor, Church, see (Church 1937b).

had been working primarily in other areas of mathematics, especially group theory. Newman's course served to focus his interests in logic; in particular, Turing became intrigued by the *Entscheidungsproblem* (decision problem) for the first order predicate calculus, and this came to dominate his thought from the summer of 1935 on. In grappling with this problem he was led to conclude that the solution must be negative; but in order to demonstrate that, he would have to give an exact mathematical analysis of the informal concept of *computability by a strict mechanical process*. This Turing achieved by mid-April 1936, when he delivered a draft of his paper, "*On Computable Numbers, with an Application to the Entscheidungsproblem*", to Newman. The "main idea" of the paper, according to Gandy, came to Turing when he was lying in Grantchester meadows in the summer of 1935. The "main idea" might have either been his analysis of computation, or his realization that there was a universal machine, and so a diagonal argument to prove unsolvability (Gandy 1988, 76). At first Newman was sceptical of Turing's analysis, thinking that nothing so straightforward in its basic conception as the Turing machines could be used to answer this outstanding problem. However, he finally satisfied himself that Turing's notion did indeed provide the most general explanation of finite mechanical process, and he encouraged the publication of this paper.

The discouraging news of Church's work reached Cambridge in May 1936. At first this seemed to pre-empt Turing's analysis of computability and his result on the Entscheidungsproblem; but Turing's definition of computability was sufficiently different from that of Church as warrant separate publication. In a letter to his mother

on 29 May 1936, Turing wrote: "Meanwhile a paper has appeared in America, written by Alonzo Church, doing the same things in a different way. Mr. Newman and I have decided however that the method is sufficiently different to warrant the publication of my paper too" (Copeland 2004, 207). Thus Turing's paper was submitted after all, on 28 May 1936; in 28 August 1936 he tacked on an appendix sketching a proof of the equivalence between his notion of computability with that of λ-definability.[170] The revised paper was published at the turn of 1936 in the *Proceedings of the London Mathematical Society*[171] and the famous "Turing machine", dubbed by Church in a review, soon become known to the whole scientific world.

We will discuss Turing's paper and his argument in the next section in detail, suffice it to say that he also proposed a version of Church's thesis, that a number is computable by finite means in the intuitive sense if and only if it can be written down by a (Turing) machine. For this claim Turing gave three types of arguments in section 9 of his paper, all elaborating the main idea that "the justification lies in the fact that the human memory is necessarily limited" (Turing 1936, 117).

3.3.3 Gödel: A Change of Attitude

In late 1935, based on a side discovery of his major result on the length of proofs

[170] The full proof came one year later while he was in Princeton working on his PhD, see (Turing 1937).

[171] There has been some confusion about the date of publication of Turing's paper, sometimes falsely cited as Turing 1937. The dating is crucial because Church, Kleene and Post all published their work in 1936 and a later date for Turing's paper would diminish his contribution in the development of the theory of computability. See Copeland's introduction to Turing's paper for a detailed explanation for possible confusion, in (Copeland 2004, 5).

where a speeding-up theorem (published as Gödel 1936) was stated, for the first time it became plausible for Gödel that, as he put it in a letter to Kreisel on May 1, 1968, that "my [incompleteness] results were valid for all formal systems." The plausibility claim relied on an observation of the "absolute" nature of computable functions in a formal system:

> It may also be shown that a function which is computable in one of the systems S_1 or even in a system of transfinite type, is already computable in S_1. Thus, the concept "computable" is in a certain definite sense "absolute", while practically all other familiar metamathematical concepts (e.g. provable, definable, etc.) depend quite essentially on the system with respect to which they are defined. (Gödel 1936, 399)

This feature of "being absolute" was again emphasized by Gödel ten years later in his contributed talk to the Princeton Bicentennial Conference on problems in mathematics about the philosophical significance of general recursiveness:

> Tarski has stressed in his lecture (and I think justly) the great importance of the concept of general recursiveness (or Turing's computability). It seems to me that this importance is largely due to the fact that with this concept one

has for the first time succeeded in giving an absolute definition of an interesting epistemological notion, i.e., one not depending on the formalism chosen[172]. In all other cases treated previously, such as demonstrability or definability, one has been able to define them only relative to a give language, and for each individual language it is clear that the one thus obtained is not the one looked for. For the concept of computability however, although it is merely a special kind of demonstrability or decidability the situation is different. By a kind of miracle it is not necessary to distinguish orders, and the diagonal procedure does not lead outside the defined notion. (Gödel 1946, 150)

In 1965 Gödel added a footnote explaining the absolute claim:

To be more precise: a function of integers is computable in any formal system containing arithmetic if and only if it is computable in arithmetic, where a function f is called computable in S if there is a computable term representing f. (Ibid.)

Following the previous letter to Kreisel about the plausibility claim, however, Gödel continued that "But I was completely convinced only by Turing's paper" (Quoted in Sieg 2006, 190). The reason for this might be that, as pointed out by Sieg,

[172] For more discussions about the importance about the notion of "formalism freeness" and Gödel's remark about the hope of finding an absolute definition of provability and definability, see (Kennedy 2013, 2017a).

(Ibid. 194-195) in order to establish that all the computable functions in any formal system containing a certain amount of arithmetic are general recursive (or Turing computable) we need to impose certain conditions on “a formal system”, and these conditions turn out to be exactly the same recursive character, thus a “subtle circularity” (Sieg 2013, 187) is committed if we want to characterize formality in terms of general recursive. We will see later that the merit of Turing’s analysis of computability lies exactly in overcoming this circularity.

The fact that in his Princeton talk Gödel mentioned “Turing computability” but only listed it parenthetically behind general recursiveness leads one writer (Ibid.) to think that at this time (1946) Gödel doesn’t distinguish the different equivalent definitions of computability and doesn’t single any of them out as having a special role, but that only in the 1951 Gibbs lecture does Turing computability becomes a focal point in Gödel’s reflections about the implication of his incompleteness theorem. This is arguably a false claim. As early as about 1938, in a lecture about undecidable Diophantine propositions, Gödel speaks too of Herbrand, Church and Turing as filling the gap for a mathematically satisfactory definition since he first published his results. After a presentation of the general recursive functions, he goes on to the remark:

> **That this really is the correct definition of mechanical computability was established beyond any doubt by Turing.** At first you can easily see that it is not possible to construct by the diagonal procedure a computable function not comprised in the definition because, although you can easily

> enumerate all possible admissible postulates, you have no procedure for deciding whether a given system of admissible postulates actually defines a function, i.e., whether it allows one to compute the values of the function. And for this reason the anti-diagonal sequence will not be computable. **But Turing has shown more**. He has shown that the computable functions defined in this way are exactly those for which you can construct a machine with a finite number of parts which will do the following thing. If you write down any number $n_1,\ldots,n_r$ on a slip of paper and put the slip into the machine and turn the crank, then after a finite number of turns the machine will stop and the value of the function for the argument $n_1,\ldots,n_r$ will be printed on the paper. (Gödel 193?, 168, my own emphasis)

The anti-diagonal feature is mentioned again in his 1946 Princeton remark quoted above as "a kind of miracle" and reminds us of Kleene's overnight conversion to Church's thesis. But this is not distinctive of Turing computability: λ-definable functions and general recursive functions too can escape the diagonalization by, so to speak, avoiding a "calculable definition of calculability" (Ibid. 167).

On numerous occasions, especially in a philosophical context, Gödel stressed again and again the importance of Turing's analysis for establishing the correct definition of a mechanical procedure or a calculable function. In his 1951 Gibbs lecture, Gödel speaks again about the superiority of Turing's definition over the others:

> The greatest improvement was made possible through the precise definition of the concept of finite procedure, … This concept, … is equivalent to the concept of a "computable function of integers" … There are several different ways of arriving at such a definition, which, however, all lead to exactly the same concept. The most satisfactory way, in my opinion, is that of reducing the concept of finite procedure to that of a machine with a finite number of parts, as has been done by the British mathematician Turing. (Gödel 1951, 304–5)

Again, in a note added in 1963 for a reprint of his 1931 paper, the first[173] published note about Turing during his lifetime, he speaks about Turing's contribution for making his theorems completely general, this time leaving out all the other definitions. Then, in his 1964 Postscriptum to his 1934 Princeton lecture Gödel wrote a verbatim assessment as 1963, but added further reasons for his appraisal of Turing. We quote it in full here:

> In consequence of later advances, in particular of the fact that, due to A. M. Turing's work, a **precise and unquestionably adequate definition** of

[173] There are in total three places in all publications during his lifetime where Gödel mentioned Turing, namely his added note to 1931, his Postscriptum to (Gödel 1934), and (Gödel 1972). Not including his conversation with Hao Wang as recorded in Wang 1974.

> the general concept of formal system can now be given, the existence of undecidable arithmetical proposition and the non-demonstrability of the consistency of a system in the same system can now be proved rigorously for **every** consistent formal system containing a certain amount of finitary number theory.
>
> Turing's work gives an analysis of the concept of "mechanical procedures" (alias "algorithm" or "computation procedure" or "finite combinatorial procedure"). This concept is shown to be equivalent with that of a "Turing machine."(Gödel 1934, 369, my own emphasis)

Gödel had an interesting note for the previous sentence: "As for previous equivalent definitions of computability, which, however, are much less suitable for our purposes, see Church (1936)… One of these definitions is given in … these lectures." (ibid.) Also, as reported by Hao Wang, Gödel is explicit that "We had not perceived the sharp concept of mechanical procedures sharply before Turing, who brought us to the right perspective" (Wang 1974, 85).

3.3.4 Why Gödel Didn't Have Church's Thesis, or Could He Have Had?

It's interesting to note first of all that Gödel doesn't mention the confluence of definitions and other quasi-empirical evidence as supporting the correctness of the

precise definitions in either the Gibbs Lecture or the 1964 postscript to the 1934 paper. The argument by confluence only makes it likely that an interesting and important class of functions has been singled out, but cannot make the further claim that this class of functions is just the effectively computably ones. Furthermore, it is possible that a systematic error has been committed. As Kreisel pointed out:

> The support for Church's Thesis… consists above all in the analysis of machine-like behavior and in a number of closure conditions, for example diagonalization… It certainly does not consist in the so-called empirical support; namely the equivalence of different characterizations: what excludes the case of a *systematic* error? (Cf. the overwhelming empirical support from ordinary mathematics for: if an arithmetic identity is provable at all, it is provable in classical first-order arithmetic; they all overlook the principle involved in, for example, consistency proofs.) (Kreisel 1965, 144)

So, apart from this insufficiency of the argument of confluence and the inductive evidence that no function has been found that is intuitively computable but not recursive, what other reasons could preventing Gödel from transforming "Gödel's Thesis" to something mathematically more definite such as Church's Thesis?

One interpretation of Gödel's failure to pursue an analysis like Church's, something Gödel definitely is in a good position to do given his depth of

understanding and experience, is given by Feferman, attributing the failure to Gödel's typical caution:

> My guess is that he also feared that no such proposal could be made convincing to the mathematical public of his day, just as the concept of truth would not be taken seriously. If so, the subsequent development showed Gödel to have been mistaken. Though certainly there were controversies about both Tarski's analysis of truth and Turing's thesis, they eventually took their place as accepted cornerstones of mathematical logic. (Solomon Feferman 1984, 163)

The main problem for Feferman's interpretation is, I think, that he overestimates the extent of Gödel's caution. This might be true for the case of incompleteness theorem, i.e., substituting the syntactical condition of omega-consistency for the semantic (and the heuristic) argument using the notion of truth, but does not apply in the case of the notion of computability. In the former case, the problem itself is a metamathematical problem about formal systems (negation-completeness) and it's desirable to only make use of restricted methods in metamathematics.[174] Furthermore, the syntactical proof of the existence of undecidable sentence G (the first incompleteness theorem) also makes the proof of the underivability of consistency of the formal system within itself (the second incompleteness theorem) much easier to

[174] At least at the time of Gödel's writing, a precise definition of the semantic notion of truth was still not available.

discover and prove. So both philosophically and mathematically Gödel's caution has good ground, which was not obviously at all, especially considering Gödel's keen attempt to make his incompleteness theorems as general as possible.

Another possible interpretation was given by Robin Gandy, who suggests that:

> But Gödel was primarily concerned—in opposition to the climate of the time—with the analysis of nonfinitist concepts and methods. This is very clearly set out in a letter of 1967 to Hao Wang (1974, 8-9). But a concern with nonfinitary reasoning is not what is needed for an analysis of calculations. Gödel admired and accepted Turing's analysis, but it is not surprising that he did not anticipate it. Indeed, to the end of his life he believed that we might be able to use nonfinitary reasoning in nonmechanical calculations. (Gandy 1988, 64)

Gandy's interpretation is also not very convincing for several reasons. First, he was careless here. What Gödel was referring to was only finite but non-mechanical reasoning not nonfinitary non-mechanical reasoning or calculation. While it's true that Gödel gave much greater credence to and had much firmer faith in nonfinitary reasoning than the formalist, he will equally embrace any definite characterization of finite concepts like computability. Actually, over the years Gödel was very interested

in characterizing the limit of finitary reasoning and constructive consistency proofs.[175]

The most plausible interpretation, as far as I can see, is given by Martin Davis. After an analysis of the situations in Princeton at that time (which we presented in more details above) Davis concluded that "thus while Gödel hung back because of his reluctance to accept the evidence for Church's thesis available in 1935 as decisive, Church (who after all was right) was willing to go ahead, and thereby to launch the field of recursive function theory" (Davis 1982, 13). With this we totally agree. However, Davis seemed to just shift the problem one step further, i.e., he didn't give an explanation to the more pressing problem of why it is that the same evidence which convinced Church didn't convince Gödel. Is it Gödel's particular philosophical stance that prevented him from launching a thesis like Church's or was Church just too hasty? Gödel was ready to use CTT as a "heuristic principle", but for him a "thesis" means much more than that. The difference can be seen in an obvious way from Post's criticism of CTT and Church's reply. Post has arrived an formulation of effective processes which he called "finite combinatorial process", the purpose of which is "not only to present a system of a certain logical potency but also, in its restricted field, of psychological fidelity" (Post 1936, 291). Although he expects the logical equivalent of his own formulation and Church's or Gödel's, he is more cautious to "offer this conclusion at the present moment as a working hypothesis" (ibid.) just like Gödel did earlier, and even if further and wider formulations maintaining more psychological fidelity turn out to be logical equivalent to the present one, this would only "change

[175] See chapter four on Gödel and Hilbert for more accounts.

this hypothesis not so much to a definition or to an axiom but to a natural law" (ibid.). In a footnote, he criticizes Church's proposal to view it as a definition:

> Actually the work already done by Church and others carries this identification considerably beyond the working hypothesis stage. But to mask this identification under a definition hides the fact that a fundamental discovery in the limitations of the mathematicizing power of Homo sapiens has been made and blinds us to the need of its continual verification. (ibid. 291)

In a review of Post's article, Church gave a reply defending himself and his choice:

> [Post] does not, however, regard his formulation as certainly to be identified with effectiveness in the ordinary sense, but takes this identification as a "working hypothesis" in need of continual verification. To this the reviewer would object that effectiveness in the ordinary sense has not been given an exact definition, and hence the working hypothesis in question has not an exact meaning. To define effectiveness as computability by an arbitrary machine, subject to restrictions of finiteness, would seem to be an adequate representation of the ordinary notion, if this is done the need for a working hypothesis disappears. (Church 1937a, 43)

Here we might have reached the core of the problem. For Church, since effectiveness in the informal sense has no exact definition it is our choice or a convention to given it a definition that is fruitful and precise while for Post and Gödel the lack of precise definition doesn't mean that the correct and adequate conceptual analysis can never be attained. A vague concept can become sharp with the right perspective and that is exactly what Gödel takes Turing's contribution to be and what is missing in Church's or his own proposed definition.[176] This realist attitude towards concepts and the optimist attitude of the possibility of an intensional conceptual analysis therewith, explains very well on the one hand Gödel's reservation about the confluence evidence that Church presented to justify his thesis and on the other hand his immediate acceptance of not only Turing machines, but more importantly the analysis Turing gave of a finite procedure. The fact that Turing machines were later proved extensionally equivalent to general recursive functions did not convince Gödel of the intrinsic merit of the other definitions. For Gödel, Turing's argument was an example of conceptual analysis, as well as a concrete application of the axiomatic method which he appreciates. We will come to see these two points in more detail when we compare Turing and Church's argument in the next section.

[176] The problem of the status of CTT will occur again later in our discussion about Gödel's criticism about Post (and Turing). Basically Gödel agrees with Post that CTT is a much more substantial claim than a definition or convention; but unlike Post who sees CTT and the related undecidability results as a "fundamental discovery in the limitations of the mathematicizing power of Homo Sapiens", Gödel sees them as a limitative result for establishing a bound, rather than the power of human reason, only for the potentiality of pure formalism in mathematics.

3.4 Assessing Church's Thesis and Turing's Thesis

In set theory there is a well-known joke alluding to the equivalency of three seemingly totally different principles (which may defy human intuition): the Axiom of Choice is obviously true, the well-ordering principle obviously false, and who can tell about Zorn's lemma? Applying to our case of computability, we have some similar: "Turing's computability is intrinsically persuasive in the sense that the ideas embodied in it directly support the thesis that the functions encompassed are all for which there are algorithms but λ-definability is not intrinsically persuasive (the thesis using it was supported not by the concept itself but rather by results established about it) and general recursiveness scarcely so (its author Gödel being at the time not at all persuaded)" (Kleene 1981b, 49). In another occasion, Kleene wrote that

> Turing's machine concepts arise from a direct effort to analyze computation procedures as we know them intuitively into elementary operations. Turing argued that repetitions of his elementary operations would suffice for any possible computation. For this reason, Turing computability suggests the thesis more immediately than the other equivalent notions… (Kleene 1967, 233)

So even Kleene, who works mainly under the tradition of λ-definability and general recursiveness[177] admits the intuitiveness of Turing's analysis. The same is true for Church. In his review of Turing's 1936 paper, he thought that Turing machine makes it "immediately clear" that Turing computability can be identified with the notion of effectiveness as it appears in mathematical contexts and made a comparison with the other equivalent notions:

> As a matter of fact, there is involved here the equivalence of three different notions: computability by a Turing machine, general recursiveness in the sense of Herbrand-Gödel-Kleene, and λ-definability in the sense of Kleene and the present reviewer. Of these, the first has the advantage of making the identification with effectiveness in the ordinary (not explicitly defined) sense evident immediately—i.e., without the necessity of proving preliminary theorems. The second and third have the advantage of suitability for embodiment in a system of symbolic logic. (Church 1937b, 43)

Curiously enough however, even though they both admitted the advantage of Turing's analysis, they never did consider Turing's argument as conceptually superior than the other characterizations, but only maybe psychologically as can be seen from the fact they never singled it out as the argument for CTT as Gödel would repeatedly

[177] Kleene worked under Church and developed natural numbers and lots of functions in λ-calculus, but turned away from it to general recursiveness in his own later (from 1936 on) important pioneering work "perhaps unduly influenced by rather chilly receptions from audiences around 1933-35 to disquisitions on λ-definability" (Kleene 1981a, 62).

do. We will discuss in detail Church's main argument and the "real stumbling block" (Sieg 1997, 165) in his analysis and finally see how Turing overcame this difficulty.

3.4.1 Church's Step-by-Step Argument and Its Flaw

In his 1936 paper Church gave a definition of the notion of an effectively calculable function of positive integers by identifying it with the notion of a recursive function of positive integers (or of a λ-definable function of positive integers), he then gave several arguments to justify his definition to such an extend that "so far as positive justification can ever be obtained for the selection of a formal definition to correspond to an intuitive notion" (Church 1936b, 100). The main argument he gave in section 7, apart from the weak confluence argument that recursive function of positive integers has the same extension as the λ-definable functions is the so-called "step-by-step argument"[178]. There are actually two approaches of the step-by-step argument, which might be called "algorithmic" and "logical" respectively:

> Thus it is shown that no more general definition of effective calculability than that proposed above can be obtained by either of two methods which naturally suggest themselves (1) by defining a function to be effectively calculable if there exists an algorithm for the calculation of its

[178] First appeared in (Gandy 1988, 72). Sieg also gave a similar argument, which might be better known, see (Sieg 1997, 165)

values (2) by defining a function *F* (of one positive integer) to be effectively calculable if, for every positive integer *m*, there exists a positive integer *n* such that *F(m)=n* is a provable theorem. (ibid. 102)

The first algorithmic approach, discussed by Schoenfield[179], is to define calculable functions directly by an algorithm. For simplicity, consider a unary calculable function F. It is reasonable to suppose that the calculation consists of writing a sequence of expressions (more precisely, without loss of generality, numerals which represent numbers) on a sheet of paper. We therefore write a_0, $a_{1, ..., a_n}$, where a_0 is initial argument m and a_n is the result of computation F(m). Now the algorithm (considered as a decision method) tells us how to derive a_n from a_0, $a_{1, ..., a_{n-1}}$ or equivalently $< a_0$, $a_{1, ..., a_{n-1}} >$. Hence there is a calculable function G such that $G(<a_0, a_{1, ..., a_{n-1}} >) = a_n$. The decision method also tells us when the computation is complete; so there is a calculable predicate P such that $P(<a_0, a_{1, ..., a_{n-1}} >)$ is false for all $i<n$ and only true for $i=n$. It is easy to notice that the attempt here to define calculability thus ends in circularity, since G and P must be assumed to be calculable. However, since G describes a single step in the calculation, it must be a very simple calculable function which can be assumed to be calculable; and the same applies to P. If we assume this, we can then prove that F is recursive.

The second "logical" approach is based on the idea of calculability in a logic. Church considers a logic L, whose language contains the equality symbol =, brackets

[179] See (Shoenfield 1967, 120). The following is a reformulation of Shoenfield argument there.

for the application of a unary function symbol to an argument, and numerals for the positive integers. For unary functions F he defines:

> F is effectively calculable if and only if there is an expression f in the logic L such that f(u)=v is a theorem of L when and only when F(m)=n is true, u and v being the expressions which stand for the positive integers m and n. (Church 1936b, 101)

He then argues that such functions F are recursive, if it's assumed that L satisfies certain conditions which are necessary if it is "to serve at all the purposes for which a system of symbolic logic is usually intended" (ibid.) These conditions are, as Church himself admitted, "substantially" the same[180] as those from Gödel's 1934 Princeton Lectures for a formal mathematical system: (1) each rule must be an effectively calculable operation, and (2) the set of axioms and rules (if infinite) must be effectively enumerable. He then interprets these to mean: (a) each rule must be recursively enumerable, (b) the set of rules and axioms must be recursively enumerable, and (c) the relation between a positive integer and the expression which stands for it must be recursive. It should be mentioned that this conception of computation as a special kind of deduction or mathematical argumentation, is indeed a very natural idea. As Kripke put it:

[180] The differences being that Gödel's conditions were formulated in the more restricted notion of primitive recursive, and Gödel used the condition that the relation of immediate consequence be recursive instead of his conditions on the rules of procedure.

> Computation is a special form of mathematical argument. One is given a set of instructions, and the steps in the computation are supposed to follow—follow deductively—from the instructions as given.
>
> In particular, the conclusion of the argument follows from the instructions as given and perhaps some well-known and not explicitly stated mathematical premises. I will assume that the computation is a deductive argument from a finite number of instructions, in analogy to Turing's emphasis on our finite capacity. It is in this sense, namely that I am regarding computation as a special form of deduction, that I am saying I am advocating a logical orientation to the problem.[181](Kripke 2013, 80–81)

The condition of recursiveness is actually sharpened in Hilbert and Bernays' 1939 *Grundlagen der Mathematik II* (Hilbert and Bernays 1939), the computable functions are also presented in terms of a logical calculus, and the conditions are reduced to a central one that the proof predicate be primitive recursive: for any m and n we can use a primitive recursive function to decide whether m codes a sequence of formulas as a proof of its last formula with a coding number n, here recursiveness is now reduced to primitive recursion. Sieg says about Hilbert and Bernays' improvement of Church's system:

[181] Kripke's point is trying to establish CTT as a corollary of Gödel's completeness theorem for first order logic, using the above idea and a weakened form of "Hilbert's thesis" that every mathematical argument can be formulated (rather than the strong form of "proved") in a first order language. See (Black 2000) for similar ideas. However, I think both arguments suffer from the same defect as Church's, as will be discussed below.

> In this way they provided the mathematical underpinnings for … Church's argument, but only relative to the recursiveness conditions: the crucial one requires the proof predicate of deductive formalism, and thus the steps informal calculations, to be primitive recursive! (Sieg 2008, 142)

Nevertheless, such a sharpening only brings the difficulty for Church to the fore in a more perspicuous way. It is indeed true that if the steps of any effective procedure (governing proofs in a system of symbolic logic) is to be recursive (which Sieg calls "Church's Central Thesis" (Sieg 1997, 165)) then any function computed in such a proof system is indeed recursive. The fatal weakness in Church's argument, however, is the core assumption that the atomic steps were stepwise recursive, something he did not justify but only taken dogmatically. Church might still take the comfort that this was not only his carelessness, but an essential circularity underneath every possible step-by-step approach, which again for Church is the most general possible definition of effective calculability. Of this point, Church is wrong, as we will see in Turing's analysis.

3.4.2 Turing's Contribution[182]

[182] For a clear and more mathematical presentation of Turing's analysis, see (Kleene 1988; Davis 1978).

It was not until the appearance of Turing's 1936 paper on computable numbers that the "stumbling block" of Church's step-by-step argument was overcome. Turing's paper, as is typical for him, was a mixture of original insights with cumbersome technicalities.[183] As Hodges once said, "the paper started attractively, but soon plunged (in typical Turing manner) into a thicket of obscure German Gothic type in order to develop his instruction table for the universal machine. The last people to give it a glance would be the applied mathematicians who had to resort to practical computation" (Hodges 1983, 124). In his brief editorial remarks for Turing's paper, Davis also warned his reader that "it may well be found most instructive to read this paper for its general sweep, ignoring the petty [and sometimes incorrect] technical details" (Davis 1965, 115). However, the ultimate correctness of Turing's results[184] and the penetrating analysis he gave for the process of computing still makes this paper stand out as uniquely important as ever.[185]

In section 9 of his 1936 paper Turing proposed three types of arguments to support the claim that his machines could compute any functions which would naturally be regarded as calculable,[186] after admitting that "all arguments which can be given are bound to be, fundamentally, appeals to intuition, and for this reason rather

[183] This maybe is due to Turing's isolation from the Princeton logicians, which turned out to have helped him. As Gandy put it: It is almost true to say that Turing succeeded in his analysis because he was not familiar with the work of others. … The bare hands, do-it-yourself approach does lead to clumsiness and error. But the way in which he uses concrete objects such as exercise book and printer's ink to illustrate and control the argument is typical of his insight and originality. Let us praise the uncluttered mind. (Gandy 1988, 83)

[184] Especially the result about the existence and construction of a universal Turing machine, whose details were corrected by Post in a careful critique, see the Appendix in (Post 1947).

[185] For an wonderful annotation of Turing's paper line by line written by a computer scientists, see (Petzold 2008).

.[186] Turing was discussing "computable numbers" rather than computable functions in his original work, that these two different ways of talking can be established equivalently easily.

unsatisfactory mathematically" (Turing 1936, 135). They are as follows:

(a) A direct appeal to intuition.

(b) A proof of the equivalence of two definitions (in case the new definition has a greater intuitive appeal).

(c) Giving examples of large classes of numbers which are computable.

Argument (c) can be seen as an empirical evidence for his claim by showing the wide range of functions (or numbers, in his context) which can be computed by his machine and doesn't differ from other similar arguments by examples and thus also suffers the same difficulty. Argument (b) is the idea of simulating first order logic system in Hilbert style by one of Turing's automatic machine, thus exhibiting the power of the machine to produce all the provable formulas. This argument, again, does not differ essentially from Church's second argument about representable functions in a logical system discussed above and shares the same weakness. The most distinctive and convincing argument of Turing is his type (a) argument: a direct intuitive conceptual analysis of what's involved in computing.

The novelty and originality in Turing's analysis of computation is that he didn't start the problem in a mathematical or logical context like Gödel or Church by asking what is a computable function, but rather attacked the problem right from its heart by considering the "real question at issue" of "what are the possible processes which can be carried out in computing a number?" (ibid.). Computing in real life is normally

done by writing certain symbols (numerals) on paper which are usually divided into squares like a child's arithmetic book. The two-dimensional character of paper, though very efficient and intuitive, is not essential to computation and thus can be replaced by an one-dimensional paper, i.e., on a tape divided into squares. Turing also suppose that the number of symbols which may be printed to be finite for "if we were to allow an infinity of symbols , then there would be symbols differing to an arbitrarily small extent"[187] (ibid.). The behavior of the computer at any moment is determined by the symbols which he is observing, and his "state of mind" at that moment. We may suppose that there is a bound B to the number of symbols or squares which the computer can observe at one moment and to observe more, he must move to other squares. The number of states of mind which need be taken into account must also be assumed to be finite, for the same reason which restrict the number of symbols, i.e., "if we admitted an infinity of states of mind, some of them will be "arbitrarily close" and will be confused" (ibid. 136). We can also assume that the type of behavior or the operations by the computer is divided into basic simple ones, that "are so elementary that it is not easy to imagine them further divided" (ibid.). Turing considered three elementary changes, including changing the symbols on the observed squares (like writing a new one or erasing the old one), moving to a new position (one square to the right or left) and a change to a new state (machine-configurations) as determined by the previous state of mind and the observed symbols. These operations seem to exhaust all the processes in a mechanical computation, and Turing thinks so, "it is my

[187] For example, we cannot tell at a glance whether 9999999999 and 999999999999 are the same.

contention that these operations include all those which are used in the computation of a number" (ibid. 118).[188] The real power of this simple and elegant model culminates in the existence of a universal machine computing any computable numbers or functions, the construction and proof of which is simply a tour de force by Turing. The accompanying analysis of the process of computing, by eliminating all irrelevant details through a sequence of simplifications and thus resulting in an imaginary machine consisting of a finite state device operating on a one dimensional infinite linear tape, is described by Martin Davis as "a remarkable piece of applied philosophy" (Davis 1982, 14). The emphasis on the symbolic process and computing operations also makes this characterization of computation independent of any formalism, being it mathematical or logical. Turing was, in a sense, creating "a shift"[189] of understanding computation. As Gandy remarked,

> All the work described above [Church, Gödel and Kleene's effort] was based on the mathematical and logical (and not on the computational) experience of the time. What Turing did, by his analysis of the process and imitations of calculations of human beings, was to clear away, with a single stroke of his broom, this dependence on contemporary experience, and produce a characterization which—within clearly perceived limits—will

[188] It is true, as Kleene later remarked about Turing's model that "the human computer is less restricted in behavior than the machine" (Kleene 1952, 377). In particular, a human being can (a) observe more than one square at a time (b) perform more complicated elementary operations (c) able to use multi rather than one-dimensional tapes and (d) choose more flexible symbolic representations than in Turing's model. However, all these more complex operations can be reduced to the atomic ones executed by a Turing machine without loss of computing power (but only efficiency or complexity) as claimed by Turing that these simple operations "include all" used in computation. For details of this reduction, see (Kleene 1952, 378–81).

[189] See (Kennedy 2017b) for further discussion about this shift of perspective.

stand for all time. (Gandy 1988, 93)

More importantly, Turing's analysis permits an axiomatic representation and, as a consequence, Turing's Thesis can be made much more precise than Church's Thesis, Gandy even claims that "it [Turing's analysis] proves a theorem" and "the proof is quite as rigorous as many accepted mathematical proofs—it is the subject matter, not the process of proof, which is unfamiliar" (ibid. 76). Let us see more closely Turing's 'proof'. Following Gandy and Sieg,[190] we shall use the term "computor" to mean an idealized human calculating in a purely mechanical/routine way, rather than "computer" which might refer to either an idealized machine (such as a Turing machine) or a physical device like a high speed digital computer. Then extracting from Turing's analysis we can get the following three conditions for the computor in an axiomatic way[191]:

(1) Boundness conditions: (B1) There is a fixed bound on the number of symbolic configurations a computor can immediately recognize; (B2) There is a fixed bound on the number of internal states a computor can be in.

(2) Determinacy condition: A computor's internal state together with the observed configuration fixed uniquely the next computation step and the next internal state.

[190] See (Gandy 1988; Sieg 1994, 1997).
[191] Our formulation follows closely Sieg's, see (Sieg 1997, 171–72) and also (Sieg 2002a, 2002b) for an extension of the axiomatic treatment of computability based on Turing's analysis.

(3) Locality conditions: (L1) A computor can change only elements of an observed symbolic configuration;

(L2) A computor can shift attention from one symbolic configuration to another one, but the new observed configuration must be within a bounded distance of the immediately previously observed configuration.

If we define computorable functions to be those that can be calculated by an idealized human computor satisfying the above axioms, then Turing actually proved that any computorable function is Turing-machine computable.[192] By this fact, Turing's Thesis then reduces to the following "central thesis" (Sieg 1997, 172):

Turing's Central Thesis: Any mechanical procedure can be carried out by a computor satisfying the above conditions.

Thus, Turing's assertion that effective calculability can be identified with machine computability is the result of a two-part analysis: the first part provides the conceptual analysis of mechanical operation by regulating several conditions (the boundness condition for symbolic configurations and number of states and the locality condition for mechanical operations) and the second part proves the mathematical fact that every

[192] Turing's computor must also satisfy the deterministic condition, however even if we allow non-deterministic, so long as they satisfy the other two conditions we get the computors having the same capacity, i.e., computing the same class of functions.

number theoretic function calculable by a computor, satisfying these conditions, is computable by a Turing machine. In such a way it is then not difficult to understand Gödel's approval of Turing's Thesis, but not Church's. Gödel favors conceptual analyses over other arguments, he sees the problem of defining computability as "an excellent example [...] of a concept which did not appear sharp to us but has become so as a result of a careful reflection" (Wang 1974, 84). The axiomatic treatment is in agreement with Gödel's earlier suggestion to Church, that "it might be possible, in terms of an effective calculability as an undefined notion, to state a set of axioms which would embody the generally accepted properties of this notion, and to do something on that basis" (Davis 1982, 9). The boundness and locality conditions can be seen essentially as properties of mechanical computation, the theorem established upon the axioms governing those properties that any computorable functions are Turing machine computable, makes Turing's Thesis beyond any reasonable doubt.

3.5 Gödel on Turing's "Philosophical Error"

3.5.1 Turing's "Philosophical Error"

Despite his unreserved appreciation of Turing's analysis for being a "precise and unquestionably adequate definition" of formal system or mechanical computability,

Gödel nevertheless published a short note in 1972 claiming to have found a "philosophical error" in Turing's argument:

> Turing in his 1937 gives an argument which is supposed to show that mental procedures cannot go beyond mechanical procedures. However, this argument is inconclusive. What Turing disregards completely is the fact that mind, in its use, is not static, but constantly developing, i.e., that we understand abstract terms more and more precisely as we go on using them, and that more and more abstract terms enter the sphere of our understanding. There may exist systematic methods of actualizing this development, which could form part of the procedure. Therefore, although at each stage the number and precision of the abstract terms at our disposal may be finite, both (and, therefore, also Turing's number of distinguishable states of mind) may converge toward infinity in the course of the application of the procedure. Note that something like this indeed seems to happen in the process of forming stronger and stronger axioms of infinity in set theory. This process, however, today is far from being sufficiently understood to form a well-defined procedure. It must be admitted that the construction of a well-defined procedure which could actually be carried out (and would yield a non-recursive number-theoretic function) would require a substantial advance in our understanding of the basic concepts of mathematics. (Gödel 1972b)

As indicated in a footnote by Gödel this remark may be regarded as a further comment on his earlier claim in 1965 that the technical results of incompleteness theorem, i.e., the existence of undecidable arithmetical propositions and the non-demonstrability of the consistency of a system in the same system for every consistent formal system containing a certain amount of finitary number theory, do not "establish any bounds for the powers of human reason, but rather for the potentialities of pure formalism in mathematics". (Gödel 1934, 370) Here Gödel was no doubt responding to the claim made earlier by Post that the generality of the incompleteness for all formal systems and the unsolvability for all methods of solvability required that CTT be seen as a "natural law", which exhibits "a fundamental discovery in the limitations of the mathematicizing power of Homo sapiens" (Post 1936, 291). But Gödel also realized that Turing's argument, as much as it establishes a correct analysis of mechanical computability, would also imply that mental procedures cannot go beyond mechanical procedures, thus placing the same kind of limitation on human reason as Post does. The puzzling problem for us is immediate: as Webb put it, how could Gödel "enjoy the generality conferred on his results by Turing's work, despite the error of its ways?". (Webb 1990, 293)

3.5.2 Resolving the Disparity

An obvious and easy way to reconcile Gödel's seemingly conflicting remarks is to distinguish different types of arguments in Turing's claim. Indeed as we discussed earlier Turing in his 1936 paper proposed three types of arguments to support the thesis that his machines could compute any functions which are calculable by finite means. Type I is a direct analysis of the operations that an ideal human computer can perform and depends on the assumption, questioned by Gödel, of finitely many "state of mind". Type II shows that the entire Hilbertian deductive apparatus of first order predicate logic can be simulated by one of his machines, i.e., a machine will produce the same theorems as the formal logic system. Type III is a "modification" of type I argument by replacing the notion of state of mind by "a more physical and definite counterpart of it", namely, a note of instructions explaining how the work is to be continued if the computer "breaks off from his work, to go away and forget all about it and later to come back and go on with it" (Turing 1936, 139). Since any stage of the computation is "completely determined" by the instructions and the symbols on the tape of a previous stage, their relation is expressible in the functional calculus, and the entire computation history could be formalized in the calculus, therefore carried out by one of his machines. That's why Turing also regard his type III argument as a "corollary" (ibid.) of II. Thus, according to Webb, Feferman maintains that "Gödel rejected only Turing's type I argument, while accepting his 'more physical' type III argument" (Webb 1990, 297). Under this interpretation Gödel rejects the type I argument because it associates the finite and mechanical nature of computational procedures with the assumption that human memory is necessarily limited, and, in particular, the number of

states of mind is bounded. He embraces the type III argument because it does not rest on this dubious assumption but on a "more physical and definite counterpart of it." Webb also suggests that Gödel was of the opinion that "all Turing was really analyzing was the concept of 'mechanical procedure,' but in his arguments for the adequacy of his analysis he overstepped himself by dragging in the mental life of a human computer" (ibid. 302). As Gödel put it, reported by Wang, that "we had not perceived the sharp concept of mechanical procedures sharply before Turing, who brought us to the right perspective" (Wang 1974, 85). The "memorial" role of mental states, that they depend on previous states and scanned symbols, is replaced in the more physical counterpart by instructions so numbered that they can refer to each other. While we may doubt the total number of the state of mind in computing, we cannot doubt the finite feature of instructions in the outer symbol space.

A similar interpretative strategy is also adopted by Shagrir, whose conclusion to Gödel's conflicting response to Turing's analysis is that "Gödel praised Turing for his analysis of an ideal human who calculates by means of finite and mechanical procedures. He was critical of what he deemed Turing's superfluous assumption that the finite and mechanical character of computation is somehow anchored in limitations on human cognitive capacities" (Shagrir 2006, 415). According to Shagrir, the finite and mechanical nature of computation is rooted in its role of defining a formal system, which is the pivotal concept in foundational debates and discussions with which Gödel was directly involved. Being finite and mechanical is exactly the defining characteristic feature of a formal system, and is neither open to question nor needs any

justification. Turing's error is to anchor these two features in the human condition, especially in the number of states of mind, i.e., Turing's has implicitly made an extra physical or materialistic assumption about mind.

The explanations of Feferman and Shagrir have the merit of being neat and are supported by textural evidence too. In a conversation with Wang, published in Wang's 1974 book, Gödel made a very similar remark with this note, but with more details. He added "Turing's argument becomes valid under two additional assumptions, which today are generally accepted, namely: 1 There is no mind separate from matter. 2. The brain functions basically like a digital computer. (2 may be replaced by: 2' The physical laws, in their observable consequences, have a finite limit of precision.)" According to Wang, however, "while Gödel thinks that 2 is very likely and 2' practically certain, he believes that 1 is a prejudice of our time, which will be disproved scientifically (perhaps by the fact that there aren't enough nerve cells to perform the observable operations of the mind)" (Wang 1974, 326). Thus Gödel apparently holds that Turing's constraints, or a version of them, apply to the brain, but not the mind. However, to begin with, to distinguish three types of arguments for Turing and then argue that Gödel agrees with one but disagree with another is too artificial. When Gödel refers to Turing's "philosophical error" he doesn't point to one particular argument but Turing's arguments as a whole. What's more, as Turing himself says, his type I argument is "only an elaboration of the ideas" presented in section 1 of his paper, and his type III argument can be regarded as a modification of type I or as a corollary of type II. So after all there is basically one central argument

based on the fact that “human memory is necessarily limited”. But not only is the above interpretation artificial, being ad hoc, but it may just be wrong. For in his argument Turing nowhere denied the possibility of the existence of an infinity number of mental states, his idea was rather that “if we admitted an infinity of states of mind, some of them will be ‘arbitrarily close’ and will be confused” (Turing 1936, 136). And the assumption made by Turing under the above restriction is that “we will also suppose that the number of states of mind **which need to be taken into account** is finite” [Ibid., my emphasis]. Even Gödel admits that this set is finite even for the mind in its current state of development, but envisaged the possibility of “systematic methods” for so actualizing the development of our understanding of abstract terms that it would “converge to infinity”.

The possibility mentioned by Gödel in his criticism of Turing was severely criticized by Kleene, one of founders of the theory of computability. In his article “Reflection on Church’s Thesis” he referred to Gödel’s contemplation as “pie in the sky” and as far as he can predict, “the pie will remain stratospheric”. The main objection of Kleene towards a potential infinity of states of mind is that “… the idea of “effective calculability” or an “algorithm” involves a set of instructions that is fixed in advance. This condition is motivated by a publicity constraint, namely, that it must be possible to “convey a complete description of the effective procedure or algorithm by a finite communication, in advance of performing computations in accordance with it. My version of the Church-Turing thesis is thus the ‘Public-Process Version’” (Kleene 1987, 493). I think Kleene’s criticism is based on some misunderstanding on Gödel’s

side. The potential infinite number of states are used by Gödel to show the possibility of an effective "mental procedure", not to deny the correctness of Turing's definition of an algorithm. It is to be noted that Gödel never used the word "effective" to describe the explicanda of Turing's analysis, and on several occasions tried hard to make this distinction. In his 1965 Postscriptum he says explicitly that "finite procedure" is equivalent to general recursiveness only if "finite" is understood to be "mechanical procedure" and the "question of whether there exist finite non-mechanical procedures not equivalent with any algorithm, has nothing whatsoever to do with the adequacy of the definition of "formal system" and of "mechanical procedure" (Gödel 1934, 370). The real problem of the dispute, I think, is Gödel's challenge to Turing's claim that allowing a finite number of states in computation does not make a serious difference "since the use of more complicated states of mind can be avoided by writing more symbols on the tape" (Turing 1936, 136). To see this point more clearly we turn back to Turing's basic fact about memory.

Webb, in his insightful introduction to this note in Gödel's collected work, also claimed that "in fact, Turing has one basic argument, which is presented in Section 1 and whose central premise is "the fact that human memory is necessarily limited". The heart of his argument was a novel abstract logical analysis of what it means to "effectively remember" things relevant to computation, such as symbols or how many times one has executed a subroutine: to do so one must be able to change from one distinguishable state to another, whether you are human or a machine. We presume indeed that states of mind may also carry memories beyond the wildest dreams of

machines, but the only ones relevant to effective computation are those you are put into by symbols and processes arising in the course of computation. But our memory is just as "necessarily limited" as a machine's—in either case, to a finite number of recognizable state changes (Webb 1990, 302). In other words, if we agree with Turing that what can be done effectively by the mind concerns memory, and since human memory is necessarily limited for otherwise some of them will be "arbitrary close" and confused, any effective procedure by the mind shall have the condition of finite number of states. However, a crucial difficulty of "second order" that might appear here is that as the number of states increases, we can effectively number all the Turing machines, and as a result a human mind becomes a "universal" mind in that it can take any machine and imitate it. It's not obvious at all that any machine can exists that is as powerful as this "universal mind". It's exactly this that shows the enormous significance of Turing's discovery of his universal machine which can take the code of any other arbitrary machine as input and imitate it on any other input. That is to say, the universal Turing machine, being infinite complex, can still exists with an internal memory of bounded complexity (finite of number of states), compensated by an external tape of symbols of unbounded size. This is exactly what Turing means when he writes that "the use of more complicated states of mind can be avoided by writing more symbols on the tape". It is right here that Gödel expresses his doubt, whether any effective state, no matter how complex, could always be compensated for in a purely symbolic way. Gödel would think that certain effective procedures involving understanding an abstract concept and its meaning would resist this symbolization,

while all formal operations "whose essence is that reasoning is completely replaced by mechanical operations on formulas" (Gödel 1934, 370) disregarding their meaning will do. That's to say, the real problem is whether in order for a state to be effectively distinguishable it has to depend purely on remembering some symbol or can it also exist in a non-mechanical way by understanding an abstract concept. I think Webb was right in seeing Gödel's criticism of Turing as based on the complexity of states rather than the potential number of states, and he is also right to point out that "it is really this kind of possibility [of a procedure involving a state exploiting meaning of abstract terms thus grasping infinitely complicated combinatorial relations] more than any convergence to an infinity of states that could undermine Turing's argument". But I think he is too hasty in saying that "once he [Turing] discovered the universal machine he saw that it could indeed compensate symbolically for a surprisingly wide class of increasingly complicated machine states" (Webb 1990, 300). As we will see, both Gödel and Turing will have more to say about what stands above Turing machines, i.e., above the mechanic method.

3.5.3 Gödel's Conception of Finite Effective yet Non-Mechanical Procedure

Gödel's idea that mental procedure, or mind, infinitely surpasses the power of any finite machine is explored in detail in his own comprehensive reflection about the philosophical significance of his own incompleteness theorems, namely, the Gibbs lecture in 1951. There the famous disjunctive conclusion which for him is an inevitable

mathematically established fact first appears: “Either mathematics is incompletable in this sense, that its evident axioms can never be comprised in a finite rule, that is to say, the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable Diophantine problems of the type specified” (Gödel 1951, 310). Although not specified in this lecture, Gödel was definitely convinced of the falsity of the second disjunct. In a lecture given around 1938, when discussing Hilbert’s famous idea that every precisely formulated mathematical question has a unique answer, i.e., a proof or disproof can always be found, Gödel points out that proof in Hilbert’s sense should be understood as “something which starts from evident axioms and proceeds by evident inferences” (Gödel 193?, 164). If, however, we formalize the notion of proof as is given in the usual logical sense as a sequence of formulas then Hilbert’s idea becomes amenable to mathematical treatment and the answer turns out to be negative, if not only because of the incompleteness theorem. However this results only shows that (1) there may exists absolutely unsolvable problems or (2) in the transition from evidence to formalism (about the notion of proof) something is lost. And Gödel was quick to point out that “it is easily seen that actually the second is the case, since the number-theoretic questions which are undecidable in a given formalism are always decidable by evident inferences not expressible in the given formalism” (ibid.) and which has the same evident character as those in the old formalism. The right conclusion to draw here is that it’s not possible to formalize mathematical evidence or mechanize mathematical reasoning even in the domain of number theory, and Hilbert’s conviction remains entirely

untouched. So the natural question is which aspects of mathematical reasoning or evidence defy formalization? In his 1974 discussion with Wang, quoted above about his criticism of Turing, Gödel mentioned two "vaguely defined" effective but non-mechanical procedures; the process of defining recursive well-orderings of integer representing larger and larger ordinals and the process of forming stronger and stronger axioms of infinity in set theory. The second procedure, of forming axioms of large cardinal numbers in set theory was Gödel's favorite example and he reiterated it again in the same 1972 note about "another version of" his first completeness theorem where he invokes an example against the idea of absolutely unsolvable problems for the human mind: "There do exist unexplored series of axioms which are analytic in the sense that they only explicate the concepts occurring in them", just like the axioms of infinity, which "only explicates the content of the general concept of set" (Gödel 1972b, 305). This emphasis on meaning and explicating abstract concepts, rather than the mere combinatorial properties of concrete symbols, lies at the heart of Gödel's conception of an effective non-mechanical procedure, which is unique for the human mind. As we noted earlier Gödel praised Turing's analysis for mechanical procedures because "this meaning [mechanical], however, is required by the concept of formal system, whose essence it is that reasoning is completely replaced by mechanical operations on formulas. Note that the question of whether there exist finite non-mechanical procedures not equivalent with any algorithm, has nothing whatsoever to do with the adequacy of [Turing's] definition of 'formal system' and of 'mechanical procedure'… (For theories and procedures in this more general sense [those involving

the use of abstract notions] the situation [with respect to complete formal systems and deciding procedures for arithmetic relations] may be different.) Note that the results do not establish any bounds for the powers of human reason, but rather for the potentialities of pure formalism in mathematics". (Gödel 1934, 370) The example Gödel gave here is his *Dialectica* Interpretation (Gödel 1958), where he used primitive recursive functionals of finite type to prove the consistency of Peano arithmetic. Gödel considered this method as "an extension of finitary method" that involves the use of abstract terms on the basis of their meaning. Abstract notions, for Gödel are "those that are essentially of second or higher order, that is, notions that do not involve properties or relations of concrete objects (for example, of combinations of signs), but that relate to mental constructs (for example proofs, meaningful statements and so on); and in the poofs we make use of insights, into these mental constructs, that spring not from the combinatorial (spatiotemporal) properties of the sign combinations representing the poofs, but only from their meaning" (Ibid.). It's exactly this characteristic ability of the human mind, the process of forming indefinitely new axioms based on the understanding and deepening of the meaning of abstract concepts that resist mechanization and formalization and make possible the existence of effective yet not mechanical procedures.[193]

[193] We cannot but hint at several central points of Gödel's discussion here. For a fuller discussion see the discussion of the next chapter on Gödel's views about the extent of finitary methods.

3.5.4 Turing and the "Mathematical Objection"[194]

The problem whether mental procedures can go beyond mechanical procedures, or whether mathematical reasoning can be mechanized, takes a more concrete form in the question "Can machines be intelligent?" in the case of Turing. In his post-war program of constructing "intelligent machines" Turing had already anticipated an objection based on the limitative logical result, which he called "the Mathematical Objection". In a 1948 report "Intelligent Machinery", Turing said that the incompleteness of Gödel and the undecidability theorem by himself and Church

> … have shown that if one tries to use machines for such purposes as determining the truth or falsity of mathematical theorems and one is not willing to tolerate an occasional wrong result, then any given machine will in some cases be unable to give an answer at all. On the other hand the human intelligence seems to be able to find methods of ever-increasing power for dealing with such problems "transcending" the methods available to machines. (Turing 1948, 410)

Take the theorem of the undecidability of first order logic as an example: Turing and Church have shown that there is no mechanical uniform method, i.e., no machine that can tell whether an arbitrary formulae of the system is a theorem or not, in some

[194] For a detailed historical account of this argument in Turing's work, see (Piccinini 2003).

cases (when given an unprovable formulae) it must fail to give an answer. On the other hand if an able mathematician is confronted with such a problem he can search around and find whatever new methods of proof available so that he is eventually able to reach a correct decision about any given formula. This possibility, even if not conclusively, should still make Gödel's claim very plausible that the human mind infinitely surpass any machine and at the same time, make Turing's conception of "intelligent machinery" very doubtful. Before we turn to Turing's reply to this seemingly very powerful mathematical objection, we will have a brief look at some of Turing's earlier considerations, especially in his conception of "ordinal logic", which gives us an impression that Turing did consider the possibility of a humanly effective yet non-mechanical procedure.

Already in his 1936 paper where he argues for the synonymity of a "machine process" and a "rule of thumb or mechanical process", he also hints at the possibility of the existence of a non-machine-computable but humanly computable number δ:

> It is (so far as we know at present) possible that any assigned number of figures of δ can be calculated, but not by a uniform process. When sufficiently many figures of δ have been calculated, an essentially new method is necessary in order to obtain more figures. (Turing 1936, 139)

This brings about the problem whether there are essential differences between the

power of human effective procedures where new mathematical methods can always be created and found while any machine or formal system, being mechanical, can only embody a fixed set of rules and methods.

In his PhD dissertation "Systems of Logic Based on Ordinals", written under the supervision of Church during 1936-1938 in Princeton, Turing took on the arduous task of overcoming Gödel's incompleteness through an infinite sequence of logical systems based on ordinals. Gödel had already proved his famous incompleteness results in 1931 that in any single logical system containing a certain amount of arithmetic like *Principia Mathematica* or other related systems there exists true but unprovable propositions. Turing's basic idea was to form stronger and stronger logical systems by adding the true yet unprovable proposition in the previous system to the next new ones as an axiom. By iterating this process infinitely many times he achieved a certain completeness result for the resulting sequence of logical systems, i.e., any true arithmetic sentences could be derived within one or another system of the sequence[195]. However, the whole sequence of logical systems was non-constructive in the sense that there exists no uniform method (or a Turing machine) that could generate all of them. Turing was fully aware that his ordinal logic wasn't a formal system in the proper sense, since the axioms and thus proofs in his sequence of systems are not decidable. In section 11 of his dissertation he explained "the purpose of ordinal logics", he introduced the interesting comparison of "intuition" and "ingenuity" in mathematical

[195] Turing's partial completeness results only apply to sentences of the same logical form as Gödel's original undecidable sentence, i.e., formulas with only universal quantifiers in front of the them. For a detailed account of Turing's work on ordinal logic and its influence for further work, see (Solomon Feferman 1988).

reasoning:

> Mathematical reasoning may be regarded rather schematically as the exercise of combination of two faculties,[196] which we may call intuition and ingenuity. The activity of the intuition consists in making spontaneous judgments which are not the result of conscious trains of reasoning. These judgments are often but by no means invariably correct (leaving aside the question what is meant by "correct"). Often it is possible to find some other way of verifying the correctness of an intuitive judgment. We may, for instance, judge that all positive integers are uniquely factorizable into primes; a detailed mathematical argument leads to the same result. This argument will also involve intuitive judgments, but they will be less open to criticism than the original judgment about factorization. (Turing 1939, 208-209)

While on the other hand, the exercise of ingenuity in mathematics was to verify the intuitive results by "suitable arrangements of propositions, and perhaps geometrical figures or drawing", i.e., to find the necessary proof for the intuitive result and to turn the intuitive judgment into a theorem by eliminating the doubt of every inferential step. Although not an absolute distinction, we can nearly always tell in particular cases

[196] In a footnote Turing said that he didn't consider the most important faculty "which distinguishes topics if interest from others", since he was here regarding the function of the mathematician as simply to determine the truth or falsity of propositions.

the different roles played by these two faculties. Through the introduction of a formal system for the corresponding mathematical field, these two faculties take an even more definite shape: intuition was reduced by setting down the axioms and formal inferential rules which are always considered to be intuitively valid while ingenuity determines which steps, among the considerable variety of possible steps in any stage of a proof, are the more profitable for the purpose of proving the desired proposition. Under this distinction, the significance of Gödel's incompleteness was even more striking:

> In pre-Gödel times it was thought by some that it would probably be possible to carry this program to such a point that all the intuitive judgments of mathematics could be replaced by a finite number of these rules. The necessity for intuition would then be entirely eliminated. (ibid. 209)

What Turing was saying, probably with Hilbert in mind, was that a single formal complete system for mathematics, where for each proposition either itself or its negation was provable, was just not possible. For the existence of such a system would eliminate intuition once and for all, leaving only the task for searching the proof, i.e., ingenuity behind. What Turing was proposing was rather the opposite by eliminating ingenuity and see "how far it is possible to eliminate intuition, and leave only ingenuity". Of course the possibility of not considering ingenuity is only conceivable in a formal system in a theoretical sense in that we can always, due to the recursive

nature of a formal system, effectively enumerate all the possible proofs and the corresponding theorems, thus replacing ingenuity with, so to speak, brutal patience. Since it is impossible to find a formal logic which wholly eliminates the necessity of using intuition, "we naturally turn to 'non-constructive' systems of logic with which not all the steps in a proof are mechanical, some being intuitive" (Turing 1939, 210). What this amounts to is that in Turing's (partial) complete system of ordinal logic some steps are non-mechanical in the sense that their validity is equivalent to the verification that some particular formulae are ordinal formulae, which again is equivalent to the truth of some number-theoretic statement. Its non-mechanical nature lies in the fact that although being true, their truth cannot be established in a fixed system with limited methods of proof, i.e., they have to be invoked sometimes as an "oracle". The wording here might suggest to some (Hodges 1988, 10) that Turing was endorsing, at least for this period of his life, an anti-mechanist view and agreeing with Gödel that mental procedures surpass the machine. However, what Turing was really stressing was only that "we want it to show quite clearly when a step makes use of intuition, and when it is purely formal. The strain put on the intuition should be a minimum" (Turing 1939, 210). That is to say, even though any single formal system would not suffice for solving arithmetic problems, a system of them might; by the same token, any single machine might not decide some propositions, many of them together might do. This fits well into the idea that although it's impossible to formalize and mechanize all mathematical intuition and its methods of proof in one single formal system, nonetheless it's possible that every mathematical intuition or method of proof

is formalisable, thus allowing more and more powerful machines approximating truth by provability as well as anyone desires.

The same strategy, by inventing stronger and stronger formal systems/machines rather than creating a machine with a faculty for intuition, is adopted again in Turing's conception of "intelligent machinery" and lies at the heart of his reply to the mathematical objection. The key assumption in the mathematical objection is that "the machine must not make mistakes. But this is not a requirement for intelligence" (Turing 1948, 411). If indeed a machine is expected to be infallible, it cannot also be intelligent, for discipline alone can never produce intelligence. Turing made the observation that

> If the untrained infant's mind is to become an intelligent one, it must acquire both discipline and initiative … But discipline is certainly not enough in itself to produce intelligence. That which is required in addition we call initiative. This statement will have to serve as a definition. Our task is to discover the nature of this residue as it occurs in man, and to try and copy it in machines. (ibid.)

What Turing writes in the rest of that paper is the possible ways to "copy" the initiative in machine, notably by "learning" through various methods of "educating" or "teaching" or "searching", or by placing a random element in machines. The result would be a machine able to alter its own instruction tables through experience and

learning, transcending the original methods available to it and thus displaying intelligence. As Copeland points out correctly: "In his post-war writing on mind and intelligence … the term "intuition" drops from view and what comes to the fore is the closely related idea of learning – in the sense of devising and discovering new methods of proof" (Copeland 2006, 168). As to the success of Turing's idea of a "learning machine", without going into any details, a quick glance of the development and extent of application of artificial intelligence today should convince ourselves of Turing's amazing insight.[197]

3.5.5 Mechanizing Mathematical Intuition: Gödel and Turing Reconciled?

Whether Turing really meant to argue for the claim that "mental procedures cannot go beyond mechanical procedures" either in 1936 or later when he was more concerned with the possibility of "learning machines", and whether Gödel's criticism and his opposing position was right or not does not seem, from our discussion above, to be as urgent a problem as it looks. What is certain is that both realized there is a "residue" part in human thinking which transcends the limitations of any particular formalism or machine, be it "intuition", "abstract understanding" or "initiative", although the way they would develop it was different. Gödel was certain about the extent of mechanical procedures with Turing machines, and explored the possibility of the existence of humanly finite effective, yet non-mechanical procedures based on an

[197] Of the few fields (chess games, language translation and learning, cryptography and mathematics) into which Turing had thought that machines could exercise their power, in most cases machines are all doing good as well, if not better, than human beings now.

understanding of abstract concepts with the help of mathematical intuition, and which also provide sufficient evidence for his conviction that human minds infinitely surpasses any finite machine. Turing, on the other hand, being more practical rather than speculative, tried to embody the "initiative" in machines too, thus making them intelligent. Whether this type of machine which can alter its own instruction table and deviate from pure discipline is still a mechanical machine in the proper sense could of course be doubted by Gödel. Nevertheless the great successes today of Turing's insight and ideas definitely prove it to be a very fruitful philosophical program[198]. The correctness of Gödel's criticism and his own philosophical position, will depend, on the one hand, on the extent of the success of Turing's program, and on the other hand the power of Gödel's own arguments and the developments of mathematics itself.[199]

[198] Much in the same status of Hilbert's program in philosophy of mathematics.

[199] Such as the possibility of deciding CH based on large cardinals in set theory, and the already successful example of proving the consistency of Peano arithmetic using the abstract notion of primitive recursive functionals in Gödel's famous *"Dialectica* interpretation".

4. Gödel versus Hilbert: Finitism and Intuition

4.1. Introduction

Despite the fact that Hilbert and Gödel never met, nor corresponded, the enormous intellectual impact of Hilbert upon Gödel, as far as problems of the foundation of mathematics and logic are concerned, is difficult not to notice. Gödel's first major logical achievement, the completeness and compactness for first-order logic as presented in his PhD thesis (Gödel 1929), is a solution to an open problem explicitly formulated in Hilbert's (and his student Ackermann's) 1928 logical book "*Principles of Mathematical Logic*". The even more celebrated incompleteness theorem was discovered by Gödel in his attempt to partially solve the major problem for Hilbert, namely, a finitary consistency proof for analysis.[200] The third great logical result of Gödel, the consistency of the axiom of choice and the generalized continuum hypothesis with the Zermelo-Fraenkel axioms, is also inspired by the methods of Hilbert in his failed attempt in his 1925 paper "On the infinite" (Hilbert 1925) due to "the great similarity in the outward structure" (Letter to Reid, in Gödel, 2003b, p.189), despite the "great differences in the heuristic ideas and the epistemological outlook" (Ibid. p.189). In contrast to this definite, direct logical and mathematical influence of Hilbert upon Gödel, the philosophical question of how Gödel really thought about Hilbert's program of consistency proof and its underlying finitism, is no doubt much more difficult and subtle. The first point

[200] "Partially" because Gödel was trying to establish the consistency of analysis relative to arithmetic first, and then a finitary proof of the consistency of arithmetic would suffice for the whole problem. For a detailed account by Gödel himself, see his 1970 letter to Yossef Balas in (Gödel 2003a, 9–10).

worth asking is what Gödel really thought about the relationship between his incompleteness theorem and Hilbert's foundational aim. Contrary to the received view that Gödel's second incompleteness about the underivability of consistency for a formal system in itself destroyed Hilbert's program, in his 1931 paper Gödel wrote only that "I wish to note expressly that Theorem XI [the second incompleteness theorem] does not contradict Hilbert's formalistic viewpoint. For this viewpoint presupposes only the existence of a consistency proof in which nothing but finitary means of proof is used, and it is conceivable that there exist finitary proofs that *cannot* be expressed in the formalism"[201] (Gödel 1931, 195). In 1958, at the latest, he has explicitly changed his mind by writing that "it is necessary to go beyond the framework of what is, in Hilbert's sense, finitary mathematics if one wants to prove the consistency of classical mathematics, or even that of classical number theory" (Gödel 1958, 241). With this change of view about the scope of finitary proof in Hilbert's sense goes along his assessment of the epistemological value of Hilbert's program and its finitary point of view in particular. In a 1938 lecture discussing the value of consistency proofs, Gödel speaks very highly of Hilbert's program: "If the original Hilbert program could have been carried out, that would have been without any doubt of enormous epistemological value" (Gödel 1938, 113). Yet in another lecture in 1961 discussing the development of the foundations of mathematics in the light of philosophy, he describes Hilbert's program as a "curious hermaphroditic thing"[202] (Gödel 1961, 379), trying to do justice both to the spirit of the time to view mathematics as a mere game with symbols with only hypothetical meaning and the nature of mathematics as a

[201] All the reference to Gödel's work, published and unpublished, will be from the five volumes of his Collected Works, (Gödel 1986; 1990; 1995; 2003a, 2003b) .
[202] The original German is "merkwürdiges Zwitterding". Martin Davis thinks "strange hybrid" would be closer to Gödel's intention, see (Davis, 2005, footnote 14).

body of truth. In a famous letter to Hao Wang in 1967 explaining the failure of Skolem to draw the conclusion of the completeness theorem for first-order logic despite all the mathematical elements in his hand, Gödel speaks of finitary reasoning in general as a "blindness (or prejudice, or whatever you may call it)" (Wang 1974, 8), and blames this prejudice for the widespread lack of the "required epistemological attitude toward metamathematics and toward non-finitary reasoning" (Ibid.). Yet, ample evidence, such as the much expanded and revised 1972 version of his 1958 *Dialectica* paper and his extensive correspondence with Bernays right until the 1970s shows that Gödel not only takes the issue of finitism seriously, but also seems a little unsettled as to the upper bound of finitary reasoning.

The above brief review naturally suggests to us three questions: a) assuming that finitary reasoning can be formalized in a comprehensive enough system, does it follow from Gödel's Second Incompleteness Theorem[203] (SIT hereafter) that consistency proof in Hilbert's sense is impossible? b) Is finitary reasoning, in the sense of both his own view and what he takes to be Hilbert's view a stable position for Gödel, or does he vacillate over time on these matters, and why? c) Is Gödel's life-long concern with finitism and constructive consistency programs coherent with his Platonism which he supposedly held since his early student days?[204]

In the next section I will first sketch an overview of Hilbert's Program (HP hereafter), distinguishing three different possible philosophical interpretations of HP and, more importantly, discuss the question whether Gödel's Theorems (both the First[205] and the Second

[203] Put it roughly, the second incompleteness theorem says that for any formal system satisfying certain modest conditions, the consistency of the system cannot be proved using methods formalizable in the system itself.
[204] In a letter to Grandjean Gödel said he "was a conceptual and mathematical realist since about 1925" (Gödel 2003a, 444).
[205] The first incompleteness theorem (FIT thereafter) says, in an informal way, that for any formal system containing some arithmetic, there exists undecidable proposition, i.e., statements which can not be proved and refuted (its negation proved).

Incompleteness Theorem) constitute a devastating blow to HP as it is originally conceived, focusing especially on the problem of the unprovability of consistency. In section three we will trace in more detail, relying on both his published papers and unpublished lectures and letters, Gödel's thought about HP in general, and finitism in particular, through the years from the publication of his incompleteness theorem in 1931 right up to 1972 when he was revising the 1958 *Dialectica* paper on an extension of finitary mathematics to prove the consistency of number theory. This careful historical study, I believe, will shed light on the question of whether Gödel's view on the nature of finitism, both his own and what he takes to be Hilbert's finitism, is stable or not. In the last section, by considering the evidence in the previous sections and by comparing the three different ways of proving the consistency of elementary number theory (the relative consistency proof through intuitionistic number theory discovered by Gödel and Gentzen independently, Gentzen's direct consistency proof with the help of transfinite induction up to ε_0[206] and Gödel's own *Dialectica* interpretation) I will give my explanation of Gödel's view of Hilbert's finitism, which differs from those offered by Martin Davis, Solomon Feferman and William Tait, and also of the related problem of the compatibility of Gödel's serious concern with finitism and constructive mathematics and his full-fledged Platonism. By focusing on the notion of "abstract intuition", I will argue that Gödel's concern with finitism and constructive consistency proof can constitute an argument for his own Platonism, albeit only in an indirect way.

[206] This is the first transfinite countable ordinal α that satisfies the equation: $\alpha = \omega^{\alpha}$.

4.2 Hilbert's Program and Gödel's Incompleteness Theorems

4.2.1 Fundamentals of HP[207]

As a matter of fact which maybe has not received enough attention, Hilbert's concern with foundational problems in mathematics started well before the set-theoretic paradoxes. The discovery of non-Euclidean geometry, the emergence of Cantorian set theory and the general movement of "arithmetization of analysis", to name just a few, are just some of the problems in the foundational debate in the second half of the 19th century, represented by Kronecker on the one side, who emphasized the fundamental importance of the integers, the constructive and decidable requirement for mathematical methods, objects and properties, and Dedekind and Cantor on the other side, who saw the nature of mathematics to be its freedom from any a priori philosophical restrictions about its methods or objects and who had no doubt about abstract or non-constructive methods as long as they were mathematically fruitful.[208] Hilbert was definitely in the Dedekind, Cantor camp, as he resisted Kronecker's tendency to restrict mathematical methods, particularly, set theory. It occurred to him that the axiomatic method was central for this debate on the nature of mathematical objects and methods. No wonder in his first foundational work, Hilbert chose to examine, as if for a test case, the axiomatic

[207] There are lots of materials written about Hilbert's program, the original one and what has become of it today. See (Zach 2003, 2006) for a wonderful exposition and (Smoryński 1988) for careful historical studies of *all* Hilbert's papers on the foundation, with special attention to Hilbert's controversy with Weyl and Brouwer.
[208] See (Avigad and Reck 2001) for a more detailed account of these two different conceptions of mathematics and their relation at that time.

method in the application of Euclidean geometry, the oldest and prestigious mathematical subject which is supposed to be free of any error. In this work, *Foundation of Geometry* (Hilbert 1899) Hilbert not only exhibited the power of formal axiomatic method but also treated in detail problems which will become central concerns in his later metamathematics, such as consistency, completeness and independence. He then proposed to use the axiomatic method, rather than the genetic one, to his treatment of the arithmetic (including integers, rational and real numbers) too, for similar reasons that he will express more forcefully 18 years later in his "Axiomatic thinking" (Hilbert 1918): "Despite the high pedagogical and heuristic value of the genetic method, for the final presentation and the complete logical grounding of our knowledge the axiomatic method deserves the first rank" (Hilbert 1900b, 1093). Accompanying the axiomatic method is the necessity of a consistency proof associated therewith, that is to say, we should never get a contradiction from the axiomatic system in any finite number of deductions, for otherwise everything can be deduced from the system and nothing at all is defined by the axioms. The set-theoretical paradoxes, especially Russell's paradoxes, however, seem to indicate a sense of uncertainty and even danger for those abstract, non-constructive reasonings. This seems to pose a real threat for Cantor's conception of mathematics as free creations, which Hilbert readily agreed with. The situation is vividly depicted by Bernays, as reported by Reid:

> Under the influence of the discovery of the antinomies in set theory, Hilbert temporarily thought that Kronecker had probably been right there. But soon he changed his mind. Now it become his goal, one might say, to do battle with

> Kronecker with his own weapons of finiteness by means of a modified conception of mathematics. (Reid 1970, 173)

This modified concept of mathematics, of course, is Hilbert's famous distinction of formal mathematics and the informal contentual metamathematics. Apart from the axiomatic mathematics strictly formalized with the tools of the new logic,[209] which was to be just "a stock of provable formulae" (Hilbert 1922, 1131), we have another mathematics—metamathematics that is supposed to safeguard the formalized one. As Hilbert put it in his first exposition of this new idea:

> … in addition to this proper mathematics, there appears a mathematics that is to some extent new, a *metamathematics* which serves to safeguard it by protecting it from the terror of unnecessary prohibitions as well as from the difficulty of paradoxes. In this metamathematics—in contrast to the purely formal modes of inference in mathematics proper—we apply contentual inference; in particular, to the proof of the consistency of the axioms. (Ibid. p.1132)

One point worth pointing out is that even if Hilbert was talking about "safeguard" or "protecting" proper mathematics, in the realm of analysis at least, he wasn't doubting the mathematical correctness of this subject, but only questioning its epistemological status, i.e.,

[209] I.e., the logical formalism from Russell and Whitehead's *Principia Mathematica*. Hilbert himself made a huge contribution too to the emergence of modern logic, see (Sieg 1999) for Hilbert's logical work before the mature HP.

only in the sense of a rigorous grounding of mathematics from our point of view. For the methods in real analysis have been pursued by mathematicians from all directions for such a long a time, and with such a thorough depth without even a shadow of inconsistency. Thus he would strongly repudiate Weyl's metaphor of "a crisis in mathematics" (Weyl 1921):

> if one speaks of a mathematical crisis, in any case one may not speak, as Weyl does, of a new crisis. He has artificially imported the vicious circle into analysis. His account of the uncertainty of the results of modern analysis does not correspond to the actual state of affairs. And as for the constructive tendencies that he and Brouwer emphasize so strongly, in my opinion it is precisely Weyl who has failed to see the path to the fulfillment of these tendencies. In my opinion, only the path taken here in pursuit of axiomatics will do full justice to the constructive tendencies, to the extent that they are natural. (Hilbert 1922, 1119)

But even Hilbert has to agree with Weyl that the "constructive tendencies" have a higher degree of evidence and intuitive certainty than the non-constructive. Errors might occur if we were to constantly and blithely apply to infinite totalities methods that are only intuitive in the finite case. For example, we are allowed to extend theorems valid for finite sums and products to infinite cases, but only under some special conditions of convergence. The general lesson to be drawn is not to restrict mathematics proper, but only work on the foundation and justification of it, i.e., metamathematics, which Hilbert considered to be the only "natural"

application of the constructive tendency:

> We therefore see that, if we wish to give a rigorous grounding of mathematics, we are not entitled to adopt as logically unproblematic the usual modes of inference [such as tertium non datur] that we find in analysis. Rather, our task is precisely to discover why and to what extent we always obtain correct results from the application of transfinite modes of inference of the sort that occur in analysis and set theory. The free use and the full mastery of the transfinite is to be achieved on the territory of the finite! (Hilbert 1923, 1140)

For Hilbert, this guarantee is achieved if we can obtain a consistency proof of the transfinite formal mathematical system, based on the intuitive reliable methods.[210] In this way, the epistemological problem of the grounding of mathematics becomes itself a mathematical question, amenable to rigorous treatment. Hilbert saw this as a sign of superiority of his method over the others in that problems about mathematics, even epistemological ones, can be dealt within mathematics itself, without bringing in an extraneous elements, such as the God of Kronecker, who gives him the notion of integers, or the special faculty of Poincaré for obtaining the truth of mathematical induction, or the primal intuition of Brouwer, or some contentual assumptions like the axiom of reducibility for the logicists Russell and Whitehead.

Another difference between Hilbert and the constructivists, apart from their view on the nature of the "crisis", and which features as the central notion of Hilbert's contributions to the

[210] For the elaboration of this idea that the consistency proof as a reliability project, see 2.23 below.

philosophy of mathematics, is his finitary point of view, which is to be the surrogate of constructive methods and which might become the be-all and end-all of his whole project:

> Kant already taught… that mathematics has at its disposal a content secured independently of all logic and hence can never be provided with a foundation by means of logic alone. … As a condition for the use of logical inferences and the performance of logical operations, something must already be given to us in our faculty of representation, certain extralogical concrete objects that are intuitively present as immediate experience prior to all thought. If logical inference is to be reliable, it must be possible to survey these objects completely in all their parts, and the fact that they occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects, as something that can neither be reduced to anything else nor needs reduction. (Hilbert 1925, 376)

As for the significance of a finitary consistency proof and what exactly 'finitary' is, we will discuss this in turn below.

4.2.2 Three Different Interpretations of HP

The central aim of HP, a finitary proof for consistency of a comprehensive enough formal

mathematical system, seems mathematically clear enough (at least as long as we leave what is exactly "finitary" behind) while the philosophical significance stays more vague. Hilbert's description and exposition of his ideas during the years seem to suggest several different interpretations and thus different possible evaluations of the eventual outcome of HP depending on the interpretation. We will give three different interpretations below and discuss its outcome in light of Gödel's FIT and SIT in the next section.

4.2.2.1 Consistency As a Condition for Mathematical Existence and Truth

From the time of his early work on geometry onwards, the question of consistency seems to have already a special importance in Hilbert's conception of the axiomatic method in general and the foundations of mathematics in particular. He shared with two of his great predecessors, namely Dedekind and Cantor, the view that mathematical activity should be free of any philosophical constraints.[211] Consistency alone is the *raison d'être*, so to speak, of mathematics. In a letter exchange with Frege about the foundations of geometry, he already expressed the view that consistency was a criterion for mathematical existence and truth.[212] More explicitly in discussing the significance of a consistency proof for the arithmetical[213] axioms in his famous Paris lecture about mathematical problems, Hilbert stressed:

[211] For a detailed description of two strands of thought for the development of mathematics in the second half of 19th century, one focusing on abstract conceptual reasoning and another on symbolic, computational manipulation, see (Avigad and Reck 2001).
[212] See Hilbert's letter to Frege on 29/December 1899, in (Frege 1982, 42).
[213] "Arithmetical" means here for Hilbert analysis, thus arithmetical axioms are a group of axioms for analysis.

> … if it can be proved that the attributes assigned to the concept can never lead to a contradiction by the application of a finite number of logical inferences, I say that the mathematical existence of the concept is thereby proved. In the case before us, where we are concerned with the axioms of real numbers in arithmetic, the proof of consistency of the axioms is at the same time the proof of the mathematical existence of the complete system of real number or of the continuum. Indeed, when the proof for the consistency of the axioms shall be fully accomplished, the doubts which have been expressed occasionally as the existence of the complete system of real numbers, will become totally groundless. (Hilbert 1900a, 1105)

In this way, as is typical of Hilbert, the philosophical question of existence is turned into a precise mathematical problem, amenable to a satisfactory solution, as is the problem of truth. In his first full description of HP in the "New Grounding of Mathematics", Hilbert is still insisting on the significance of consistency for mathematical truth:

> Accordingly, a satisfactory conclusion to the research into these foundations can only be attained by the solution of the problem of the consistency of the axioms of analysis. If we can produce this proof, then we can say that mathematical statements are in fact incontestable and ultimate truths—a piece of knowledge that (also because of its general philosophical character) is of the greatest significance for us. (Hilbert 1922, 1121)

How does consistency, a merely syntactical notion about symbols, formulas and mechanical notion of derivability lead to the substantive notion of existence and truth? We will discuss in the next section the plausibility of this thesis in light of Gödel's theorem and the related question of whether Hilbert can be said to be a formalist under such an interpretation of HP.

4.2.2.2 HP As a Conservative Program

Unlike the first interpretation of the significance of consistency proofs for such a sweeping aim of existence and truth for mathematics, this view focuses more on the distinction of finitary/real/contentual and transfinite/ideal/formal element in mathematics and refrains from considering any questions about truth and existence in the ideal part by showing that it is dispensable, or put in another way, the idea is reducible to the real. The ideal part of mathematics is conservative over the real part if any real statement proved with the help of ideal elements can also be proved by real, finitary means alone. If, however the ideal part is not conservative over the real part, then it will prove a real statement P not provable by finitary means. Assuming that all finitary truths are provable, then P must be a provably false real statement. Together with its negation this will make the whole theory inconsistent. So, consistency guarantees the conservativity of the ideal part over the real part of a formalized mathematical system. The use of the ideal elements in mathematics is just for pragmatic reasons: it is easier to find the proof, simplify some complex proofs or unite different proofs

into a single one etc..[214] Even though Hilbert never clearly articulated conservativity of the ideal over the real for finitary (especially finitary general) statements as an aim of his foundational project, several passages of his exposition of HP suggests such a reading, for example:

> In my proof theory we adjoin to the finite axioms the transfinite axiom and formulae, just as one introduces imaginary elements to the reals in the theory of complex numbers and ideal objections in geometry. And the motive for doing this and the success of the procedure is in my proof theory the same as there: namely, the addition of the transfinite axiom achieves in a sense the simplification and completion of the theory. (Hilbert 1923, 1144)

And also in his 1927 Hamburg lecture:

> But even if one were not satisfied with consistency and had further scruples, he would at least have to acknowledge the significance of the consistency proof as a general method of obtaining finitary proofs from proofs of general theorems—say

[214] The conservativity reading of HP sometimes goes hand in hand with an instrumental reading as well. But "instrumental" itself is not sharp enough to determine whether the consequences should be conservative or not. A telescope is instrumental in the sense of allowing us see more things which are not possible without this instrument, while a car has an instrumental value compared with walking by foot in the sense that it's more efficient, but not, in principle more fruitful in terms of the final result.

of the character of Fermant's theorem—that are carried out by means of the ε-function. (Hilbert 1927, 374)

4.2.2.3 HP As a Reliability Program

On my view, the most plausible reading of HP in its mature form is to see it as a kind of reliability program. According to HP, finitary reasoning has the advantage over the transfinite in that the intuitive evidence is lost in the latter while certainty and reliability are preserved only in the former. It is then natural to extend this certainty and reliability in the finitary to the transfinite, even if only in a derivative sense by some sort of method guaranteed by mathematics itself. This is exactly what Hilbert had in mind when he pronounced that:

> My investigations in the new grounding of mathematics have as their goal nothing less than this: to eliminate, once and for all, the general doubt about the reliability of mathematical inference. (Hilbert 1923, 1136)

Compared with the conservativity interpretation, we no longer restrict *a priori* the role transfinite reasoning might play, that is to say, whether it will lead exactly to what only finitary reasoning can prove, but we only require what is proved by transfinite reasoning to be reliable and contentually true. And this is to be achieved by mathematical rigour:

> Rather, our task is precisely to discover why and to what extent we always obtain correct results from the application of transfinite modes of inference of the sort that occur in analysis and set theory. The free use and the full mastery of the transfinite is to be achieved on the territory of the finite! (Hilbert 1923, 1140)

Now, assume that we include universal statements[215] in Hilbert's finitary part of mathematics, then the reliability interpretation amounts to saying that all provable Π_1 sentences are true, which, again is trivially equivalent to their consistency.[216] Put it in another way facilitating our discussion below, a more formal presentation of Π_1-soundness is the so-called Π_1-reflection principle for a system T, i.e., for any Π_1 sentence φ, $\mathrm{Prov}([\varphi]) \rightarrow \varphi$, (where Prov(x) is the formal provability predicate and $[\varphi]$ is the Gödel number of φ in any reasonable numbering), is itself provable from T.

4.2.3 Gödel's Theorems and Their Relevance

In this part I will discuss in turn whether any of the three interpretations of HP is tenable, in view of Gödel's theorems, both the first and the second incompleteness theorem. Our main conclusion is that, no matter which way you take HP to be, its main claim cannot stand against

[215] These are what logicians call Π_1 sentences, sentences starting with a (or a sequence of) universal quantifier and whose main part are quantifier-free. Some universal formulas must belong to the finitary general sentences since the statement asserting the consistency of a formal system T is of such a form.
[216] If some provable universal statement is false, then the counterexample, being a Δ_0 sentence, is provable. Its negation follows also from the universal statement by instantiation, thus a contradiction and inconsistency.

Gödel's theorem.

4.2.3.1 Consistency and Mathematical Truth and Existence

The question whether consistency guarantees existence and truth receives a more definite form if we explain these notions as existence and truth *in a model*. The original philosophical claim then becomes that every consistent (set of) sentence has a model. This, again, is equivalent to the claim that a formal system is complete in the sense that every valid sentence, i.e., sentence true in every model can be proved. First-order logic is just such a system. However, Gödel's first incompleteness theorem already casts a doubt on such a general claim. In the case of second-order arithmetic T, for example, suppose G is an undecidable proposition, then both G and the negation of G are consistent with T, and will each have a model, which will be in direct contradiction with the well-known fact that second-order arithmetic is categorical, namely all its models are isomorphic.

If, however, we don't understand existence and truth in a model-theoretical way, then what sense can we make out of this formalistic way of speaking of existence and truth?

First it is to be noted that Hilbert definitely wasn't a formalist in the sense of viewing mathematics as a pure game with an inventory of meaningless formulas.[217] True, there are passages where he mentioned transfinite propositions as devoid of meaning, but that's more

[217] For example, Ramsey characterized Hilbert's ideas to regard mathematics "as a sort of game, played with meaningless marks on paper rather like noughts and crosses" (Ramsey 1926, 188) according to fixed rule.

likely to be a consequence of his finitary view to giving meaning to only what is directly intuitive and therefore cannot be taken at its face value. Rather, lots of evidence weighs against such a simple formalist interpretation. We only need to mention his passionate commitment towards mathematics when he describes it to be "our most valuable treasures" (Hilbert 1922, 1119) and compares analysis to be "a single symphony of the infinite" (Hilbert 1925, 373) and more famously Cantor's set theory as a "paradise" out of which we should never be driven out. Moreover, Hilbert never saw his formalization as a formula game, but as having "besides its mathematical value, an important general philosophical significance. For this formula is carried out according to certain definite rules, in which the technique of our thinking is expressed. These rules form a closed system that can be discovered and definitely stated. The fundamental idea of my proof theory is none other than to describe the activity of our understanding, to make a protocol of the rules according to which our thinking actually proceeds". (Hilbert 1927, 475)

Weyl, in criticizing the view of transforming mathematics into a meaningless formula game often attributed to Hilbert, suggests that:

> If mathematics is to remain a serious cultural concern, then some sense must be attached to Hilbert's game of formulae, and I see only one possibility of attributing to it (including its transfinite components) an independent intellectual meaning. In theoretical physics we have before us the great example of a [kind of] knowledge of

> completely different character than the common or phenomenal knowledge that expresses purely what is given in intuition. While in this case every judgment has its own sense that is completely realizable within intuition, this is by no means the case for the statements of theoretical physics. In that case it is rather the system as a whole that is in question if confronted with experience. (Weyl 1925, 140)

The analogy with physics is striking and actually Hilbert himself used a similar analogy in 1928. He there suggested that we don't need to give an interpretation to every individual sentence by itself for "a theory by its very nature is such that we do not need to fall back upon intuition or meaning in the midst of some argument" (Hilbert 1927, 457). Just as the physicist demands that only some observational sentences are derived from laws of nature or hypotheses solely by inferences just like a formula game without extraneous considerations being adduced, in Hilbert's proof theory, only finitary propositions deduced from transfinite part need to be "directly capable of verification" (Ibid., 457). In such a way, mathematics is turned to a theoretic science just like physics. The analogy with physics is therefore not that transfinite propositions have no meaning just as propositions involving theoretical terms have no meaning, but that transfinite propositions require no *direct intuitive meaning* just as one does not have to directly see electrons or fields in order to theorize about them. Howard Stein summarizes this point in a very impressive way:

> [Hilbert's] point is, I think, this rather: that the mathematical *logos* has no responsibility to any imposed *standard* of meaning: not to a Kantian or Brouwerian "intuition", not to finite or effective decidability, not to anyone's metaphysical standards for "ontology"; its *sole* "formal" or "legal" responsibility is to be consistent (of course, it has also what one might call a 'moral' or 'aesthetic' responsibility: to be useful, or interesting, or beautiful; but to this it cannot be constrained – poetry is not produced through censorship). (Stein, 1988, p. 255)

With this I very much agree, should the consistency program succeed. We can then take the transfinite propositions either as if they are as meaningful as the finite, or if we like, take them as meaningless formula to be played in a symbol game. However, no matter how meaning is understood as the normal "model-theoretic" or, as Weyl suggests, "an independent intellectual" way, its underlying assumption, consistency, seems to pose an even harder problem than the choice of meaning which comes from consistency.

4.2.3.2 Conservativeness and Inexhaustibility

The interpretation of HP on the pursuit of consistency as leading to the conservativity of a transfinite theory T over its finitary part S, discussed above in 2.22, fails in a most striking way due to Gödel's FIT. That is to say, given any formalization of S, there exists undecidable

propositions in S but which will become decidable with the help of T. (Those undecidable propositions, under certain syntactical restrictions, will usually turn out to be true as well). The consistency of T itself, for example, will become such a proposition. This contrasting situation is famously illustrated by what Gödel called the "inexhaustibility of mathematics". Already in 1933 Gödel noticed this "strange situation" whereby we set out to find a formal system for mathematics but instead end up finding an infinity of systems always open to extension (either type theory or set theory). Instead of viewing this situation as unsatisfactory or as discrediting the theory of types, Gödel saw it as "in perfect accord with" (Gödel 1933c, 48) his FIT that even for the theory of integers new methods of proof or axioms will always be needed.

> This fact is interesting also from another point of view, it shows that the construction of higher and higher types is by no means idle, but is necessary for proving theorems even of a relatively simple structure, namely arithmetic propositions… The theorem of Goldbach, which states that each even number is the sum of two prime numbers; would be an example of an arithmetic proposition in this sense. A special case of the general theorem about the existence of undecidable proposition in any formal system is that these are arithmetic propositions which can be proved only by analytical methods and, further, that there are arithmetic propositions which cannot be proved even by analysis but only by methods involving extremely large infinite cardinals and similar things. (Ibid. p. 48)

4.2.3.3 Provable Consistency Statements and Reflection Principle

There are claims against the view that Gödel's second incompleteness of the underivability of the consistency of a formal system within itself makes a finitary consistency proof impossible (even assume the formal system is comprehensive enough to include all finitary means of proof) by arguing from the possibility that some consistency statement *Con* is actually provable from the system[218]. In fact, there do exist some non-standard formal systems of arithmetic whose consistency is guaranteed simply by the way they are presented. Moreover, for any standard formal system of arithmetic, we can indeed find a consistency-guaranteed system with the very same theorems, assuming that the standard system is consistent. However these systems fail to be genuine formal mathematical systems for one reason or another and therefore do not provide a real escape for HP under the second incompleteness theorem.[219] On the other hand, similar to those "consistency-minded" formal systems, there also exists "non-standard" consistency statements built up from a "consistency-minded" proof predicate which are provable in the same standard formal system of arithmetic T. Suppose Prf(x, y) is the standard formal proof predicate for a certain formal system containing arithmetic, i.e., this arithmetized predicate represents the metamathematical notion that x is the Gödel number of a proof consisting of a sequences of formulas with the end formula having Gödel number y. Then the canonical consistency statement *Con* is usually defined to be $\forall x(\neg \mathrm{Prf}(x, [\mathbf{0} = \boldsymbol{s}\mathbf{0}]))$, where $[\mathbf{0} = \boldsymbol{s}\mathbf{0}]$ is the Gödel number of the formula whose arithmetic interpretation is that 0 equals 1, i.e., a contradiction. Assuming T contains a

[218] This phenomena was first systematically studied by Feferman, see (Solomon Feferman 1960).
[219] For a discussion of examples of this kind, see (Giaquinto 2002, 188) where he discusses the inadequacy of Feferman and Rosser's system. Their systems fail to be genuine formal mathematical system because the proof relation is not decidable and the cut rule doesn't hold, respectively.

certain sort of induction principle, we can indeed show that this canonical consistency statement cannot be provable from T itself.[220] Now instead of the stand proof predicate, we can indeed introduce a new "consistency-minded" proof predicate associated with it. As before, let Prf(x,y) express T's derivability relation, now define:

MPrf(x,y)$\overset{\text{def}}{=}$Prf(x,y) $\wedge$ $\neg$Prf(x, $[\mathbf{0}=\mathbf{s0}]$)[221]. The intuitive meaning for this new proof predicate is that x is to be (the Gödel number of) a M-proof of y if and only if x is a proof with end formula y and x is not a proof of contradiction, 0=1. Suppose that T is consistent, then any proof sequence of formulas cannot end with a contradiction, that is to say, $\neg$Prf(x, $[\mathbf{0}=\mathbf{s0}]$ is true for all number x and thus derivable (being a quantifier-free sentence, i.e., Δ_0 sentence). Hence, from an extensional point of view, MPrf(x, y) and Prf(x,y) are exactly true of the same ordered pair of numbers as their arguments if T is consistent. They differ, however in their "intensions" in that MPrf(x,y) says much more than Prf(x,y). Now if we construct a consistency statement out of this "consistency-minded" proof relation by defining Con*$\overset{\text{def}}{=}$ $\forall$x($\neg$MPrf(x, $[\mathbf{0}=\boldsymbol{s0}]$)), then Con* is easily derivable from T since it is a theorem of predicate logic, i.e., the law of non-contradiction:

Con*$\leftrightarrow$ $\forall$x$\neg$[(Prf(x, $[\mathbf{0}=\boldsymbol{s0}]$) $\wedge$ $\neg$Prf(x, $[\mathbf{0}=\mathbf{s0}]$)].

Considerations like the above lead some commentators, for example Detlefsen (see Detlefsen, 1986) to doubt the effect of Gödel's second incompleteness on HP. However, closer examinations will show, I believe, that the argument involving the provability of certain "consistency statements" and similar statements, no matter how significant they are in

[220] For a relatively detailed proof, see (Smith, 2013, chapter 31).
[221] This neat example comes from Mostowski, that's the reason why we use MPrf in the definition, see (Mostowski 1966, 24).

revealing the difference between the two incompleteness theorems, cannot do the philosophical job they are supposed to for the following three reasons.

To start with, provable consistency statements involving ways of "consistency-minded" provability predicates, like the one above or any other, do not really have a philosophical interest for the simple reason that whether those statements genuinely asserting consistency depend again on whether the theory in question is consistent itself. If T is indeed consistent, then the provability predicate MPrf(x,y) using a little trick does express the genuine proof relation and is co-extensive with Prf(x,y), and the related Con* says that T is consistent. But, if T isn't consistent, then MPrf doesn't express the orthodox proof relation and Con* cannot be read as expressing normal consistency. In the case above, suppose that T is inconsistent and m is the Gödel number of a proof with theorem $\mathbf{0} = \mathbf{s0}$. Then we have both Prf(m, [$\mathbf{0} = \mathbf{s0}$]) and ¬MPrf(m, [$\mathbf{0} = \mathbf{s0}$]), but everything is provable in an inconsistent system. ¬MPrf(m, [$\mathbf{0} = \mathbf{s0}$]) only shows that MPrf(x,y) and the related Con* is not really the one we are after.

Secondly, not unrelated to the first point is the fact that usually the "consistency-minded" provability predicates cannot satisfy all three Hilbert-Bernays-Löb (HBL) derivability conditions,[222] which can be considered to be just reflections of our intuition about what a proof predicate should be and which are also essential to guarantee the unprovability of the consistency statement in the system itself associated with such a standard proof predicate[223]. The three HBL conditions for the formal theorem predicate "Prov(x)" (Prov(x)$\stackrel{\text{def}}{=} \exists y \text{Prf}(y, x)$) for any formal mathematical system T are that:

[222] These conditions are picked out in (Hilbert and Bernays 1939) where they presented the first full proof of Gödel's second incompleteness theorem, and then much simplified by Löb in (Löb 1955).
[223] It's true that Jeroslow had a version to prove the second incompleteness without relying on condition (3) above. See (Jeroslow 1973). However I think it's more of a technical convenience rather than any philosophical challenge for the discussion below.

(1) If A is a theorem of T, and [A] is the numeral denoting the Gödel number of A , then Prov([A]) is also a theorem of T.

(2) For any formula A of T, it is a theorem of T that "If Prov([A]), then Prov([Prov([X])])". That is to say, condition (1) is itself formalizable in T, i.e., T "knows" it too.

(3) For any formula A and B of T, the following is provable: "If Prov([A]) and Prov([A→B]), then Prov([B])". This amounts to saying that the inference rule modus ponens is formalizable in T itself.

Any non-standard proof predicate will fail to satisfy these conditions in one way or another, thus making it possible for the associated consistency statements to be provable in the same system, apparently violating Gödel's second incompleteness theorem. Consider another example, the famous Rosser predicate can be defined as follows:

RPrf(x,y)$\stackrel{\text{def}}{\equiv}$ $\mathrm{Prf}(x, y) \wedge \forall z \leq x \neg \mathrm{Prf}(z, \mathrm{Neg}(y))$, where Neg(y) is the Gödel number of the negation of the formula whose Gödel number is y.

Again, when the system T is consistent, then RPrf(x,y) has the same extension as Prf(x,y), but when T is inconsistent, it turns out that some formula might not be "provable" in the sense of Rosser provability. It can be shown that Rosser Provability fails to satisfy all the three conditions above. The primary interest in Rosser proof predicate lies in the fact that it seems to capture more closely the way mathematicians deal with mathematical proofs. Whenever they have a proof of some theorem, they want to make sure that the opposite has never been proven "before", in some predetermined sense. However, no matter how plausible it looks to earn its

merit as a model of actual mathematical practice, even disregarding the practical problem of ordering proof in some feasible way, it fails to provide an escape route for HP for the reason that it is exactly HP that it is trying to salvage. For the essence of proof in HP is those formalized "ideal" reasoning, and thus a proof must be understood in the formal sense too, i.e., a sequence of formulas starting with axioms and complying with inference rules each step with the end formula being proved as the theorem, without bringing in any extraneous element. In the particular case of the Rosser proof predicate, if T is inconsistent, then some false theorems in the real part of HP might still be Rosser provable: the counterexample might appear much later in the ordering of the proof, which would be a major violation of HP as a reliability program.

Lastly and maybe most importantly, we can sidestep the question of finding the "right" provability predicate and thus the "right" consistency statements in order to show the inadequacy of HP. In a 1972 note called "The best and most general version of the unprovability of consistency in the same system" (Gödel 1972b, 305), Gödel distinguishes two senses of consistency: "inner consistency" in the sense of non-demonstrability of both a proposition and its negation and "outer consistency", in the sense of "the rules of the equational calculus applied to equations demonstrable in s between primitive recursive terms yield only correct numerical equations" (Ibid. p. 305). In the context of our discussion here, inner consistency corresponds to the standard consistency statement *Con* for a system T, while outer consistency corresponds to the Π_1 reflection principle mentioned earlier as the most suitable interpretation of HP as a reliability program. Now, for usual systems (systems whose

consistency statement comes from a proof predicate satisfying the HBL condition) these two senses of consistency is equivalent[224]. However, even if the HBL conditions fail for the proof predicate, the Π_1 reflection principle is still underivable in T itself assuming that T is in fact consistency, in contrast to those non-standard, consistency-minded consistency statements. The reason for this is just an easy consequence of the first part of Gödel's FIT. Take a canonical sentence G as an undecidable sentence for T, then since $G \leftrightarrow \neg Prov([G])$, being consistent is enough a condition for G to be underivable in T.[225] Since G is itself a Π_1 sentence, it follows that if Π_1 reflection principle holds for T, then $Prov([G]) \rightarrow G$ should also be provable in T. This, together with $G \leftrightarrow \neg Prov([G])$ will make G provable in T, contradicting the assumption that G is not provable in T. Thus, consistency together with some very generous conditions (such as the provability of the diagonalization lemma[226]) already makes Π_1 reflection principle an unachievable goal for any system T, but Π_1 reflection principle is just what is needed "in order to 'justify' the transfinite axioms of a system S in the sense of Hilbert's program". (Ibid. 305)

4.3 Gödel's Discussion of HP and Finitism in General

In this section I will mainly be gathering evidence on Gödel's views on HP and finitism from his publications and several important unpublished lectures and correspondence, in a

[224] See (Smith, 2013, section 36) for a proof of this fact.
[225] While ω-consistency is needed to show that the negation of G is also underivable, thus making G undecidable in Gödel's original (and most natural) proof.
[226] This is the lemma which says that for any open formula $\Phi(x)$ expressible in the language, there exist a sentence P such that $P \leftrightarrow \Phi([P])$ is provable in the formal system.

roughly chronological order, paying special attention to their evolution and differences.

4.3.1 Gödel's Caution in 1931:

Gödel in his watershed 1931 paper on formally undecidable propositions left open the possibility that there could be finitary methods which are yet not formalizable in the systems under discussion, thus his theorem doesn't constitute a conclusive argument against HP:

> "I wish to note expressly that Theorem XI (and the corresponding results for M and A)[227] do not contradict Hilbert's formalistic viewpoint. For this viewpoint presupposes only the existence of a consistency proof in which nothing but finitary means of proof is used, and it is conceivable that there exist finitary proofs that cannot be expressed in the formalism of P (or of M or A)." (Gödel 1931, 195)

It is interesting here to compare Gödel's cautious attitude with the other two great logicians in Hilbert's school at the time, namely, von Neumann and Bernays[228]. Being the first person to hear from Gödel's announcement of his first incompleteness theorem in the 1930 Königsberg meeting and understand it[229], von Neumann is also quick enough to discover the second incompleteness theorem about the unprovability of consistency, independently of Gödel too.[230]

---

[227] Theorem XI is the Second Underivability Theorem for a version of *Principia Mathematica* (with reducibility), and M and A are formal systems of set theory and analysis respectively.

[228] Herbrand, another important member of the Hilbert school, has a similar opinion as von Neumann, writing to Gödel that "it is impossible to prove that every intuitionistic proof is formalizable in Russell's system, but that a counterexample will never be found. There we shall perhaps be compelled to adopt a kind of logical postulate". See (Sieg 2005) for a comprehensive account about the correspondence of Herbrand with Gödel.

[229] See (Wang 1981, 654–55) for an account of von Neumann's encounter with Gödel in the meeting.

[230] Gödel's 1931 paper was received on November 17, 1930 while von Neumann's letter containing his discovery of

However, he disagrees strongly with Gödel about the effect of this theorem on HP. In a letter of 29 November from him to Gödel, he writes:

> I believe that every intuitionistic consideration can be formally copied, because the "arbitrarily nested" recursions of Bernays-Hilbert are equivalent to ordinary transfinite recursions up to appropriate ordinals of the second number class. This is a process that can be formally captured, unless there is an intuitionistically definable ordinal of the second number class that could not be defined formally—which is in my view unthinkable. Intuitionism clearly has no finite axiom system, but that does not prevent its being a part of classical mathematics that does have one.
>
> Thus, I think that your result has solved negatively the foundational question: there is no rigorous justification for classical mathematics.

In another letter, after he received the galleys of Gödel's 1931 paper where Gödel expressed the above open possibility of finitary consistency proofs, von Neumann replied this time in a more forceful way:

---

the second incompleteness theorem arrived to Gödel on November 20, 1930, only 3 days later! It must be such a relief for Gödel to have his name on both of the celebrated incompleteness theorem!

I absolutely disagree with your view on the formalizability of intuitionism. Certainly, for every formal system there is, as you proved, another formal one that is (already in arithmetic and the lower functional calculus) stronger. But intuitionism is not affected by that at all. (Gödel 2003b, 341)

And he further explains that:

Clearly I cannot prove that every intuitionistically correct construction of arithmetic is formalizable in A or M or even in Z—for intuitionism is undefined and undefinable. But is it not a fact, that not a single construction of the kind mentioned is known that cannot be formalized in A, and that no living logician is in the position of naming such [a construction]? Or am I wrong, and you know an effective intuitionistic arithmetic construction whose formalization in A creates difficulties? If that, to my utmost surprise, should be the case, then the formalization should certainly work in M or Z!

I would be very grateful if you would tell me whether you are really conjecturing the existence of such examples, or whether you even know some? (Ibid, p.343)

Concerning the use of "intuitionistic consideration" or "intuitionism" it should be noted

that the prevailing view in the Hilbert school, also including Gödel in the early 1930s, is to equate finitism with intuitionism, according to Bernays, (1967, 502). The surprising discrepancy between these two notions only comes with the relative consistency proof of Peano Arithmetic (PA) over Heyting Arithmetic (HA) which is a formalization of intuitionistic arithmetic, discovered independently by Gödel and Gentzen in 1933 (Gödel 1933b; Gentzen 1933). Gödel's reservation about the scope of finitism, or what comes to the same, intuitionism seems to lie in the fact that the totality of all intuitionistically correct proofs cannot be contained in one single formal system. This is reinforced by evidence from Gödel's correspondence with Bernays at nearly the same time. In a letter from Bernays to Gödel on 18th of January 1931, he first expressed his agreement with Gödel about the impact of his theorem on finitary consistency proofs: "if, as von Neumann does, one takes it as certain that any and every finitary consideration may be formalized within the system P — like you, I regard that in no way as settled — one comes to the conclusion that a finitary demonstration of the consistency of P is impossible" (Gödel 2003a, 87). In the same letter he also discussed the problem of the addition of a kind of ω-rule for the completeness of a system, already introduced by Hilbert in his (Hilbert 1931a)[231]. Roughly speaking, the rule allows one to deduce the conclusion $\forall x A(x)$ (A being quantifier-free) for which it has been shown finitarily that each instance A(n) for any natural number n is already provable in Z*. Gödel's reply, apart from pointing out that even with this new rule a system is still incomplete, stressed that:

[231] Whether Hilbert proposed this rule out of the natural development of his project or out of his reaction to Gödel's incompleteness theorem, evidence seems to be not conclusive, at least as far as what we can gather from (Bernays 1935a) and (Reid 1970, 198–99).

> By the way, I don't think that one can rest content with the systems [Z^*, or Z^{**}] as a satisfactory foundation of number theory (even apart from their lack of deductive closure), and indeed, above all because in them the complicated and problematical concept "finitary proof" is assumed (in the statement of the rule for axioms) without having been made mathematically precise. (Gödel 2003a, 97)

4.3.2 The Cambridge Lecture in 1933:

However, what for Gödel was "complicated and problematical" in 1931 seems to take a more definite shape only two years later, when he gave a lecture entitled "The present situation in the foundation of mathematics" for a meeting of the Mathematical Association of American on December 30, 1933. This is the first of a series of remarkable lectures that he gave (Gödel, 1933/1938/1941) in which he discussed, more than in any other published or unpublished papers, ideas about finitism and Hilbert's program. Gödel begins this "rich and, in certain respects, remarkable article" (Solomon Feferman 1995, 36) by stating the two-part (mathematical-philosophical) aim of the foundation of mathematics to be that of reducing the methods of proof by mathematicians to a minimum number of axioms and primitive rules of inference and some justification for those axioms, i.e., "a theoretical foundation of the fact that they lead to results agreeing with each other and with empirical facts" (Gödel 1933c, 45). The first aim was solved "in a perfectly satisfactory way" (Ibid., 45) by means of the theory of types and by axiomatic set theory as well, which is "nothing else but a natural generalization

of the theory of types, or rather, it is what becomes of the theory of types if certain superfluous restrictions are removed" (Ibid., 46). The second aim however, according to Gödel, was "extremely unsatisfactory" for the three difficulties associated with the meaning of the formalism, namely, the non-constructive notion of existence, the impredicative definition of classes/properties and the axiom of choice. Justifications for problems like this led Gödel to the stunning remark that "our axioms, if interpreted as meaningful statements, necessarily presuppose a kind of Platonism, which cannot satisfy any critical mind and which does not even produce the conviction that they are consistent" (Ibid. p. 50).[232] Even though we have abundant inductive evidence that the axioms haven't led to any contradiction, we still might want to have a formal proof of this conviction that they will never lead to any contradiction, and this naturally leads to consistency program proposed by Hilbert. However, the consistency proof must be conducted in such as a way that is unobjectionable, i.e., "it must strictly avoid the non-constructive existence proofs, non-predicative definitions and similar things, for it is exactly a justification for these doubtful methods which we are seeking" (Ibid.,50). But mathematics free of those problematic methods, what Gödel still calls "intuitionistic mathematics", is not uniquely determined as it might first seems to so, but that "it is certainly true that there are different notions of constructivity and, accordingly, different layers of intuitionistic or constructive mathematics. As we ascend in the series of these layers, we are drawing nearer to ordinary non-constructive mathematics, and at the same time the methods of proof and construction which we admit are becoming less satisfactory and less convincing"

[232] This particular statement has been used by commentators like Martin Davis to doubt the authenticity of Gödel's own claim that he has held a platonistic view since 1925, and he pointed out further that "it seems clear that in 1933 Gödel's beliefs were quite different from those of his late years" (Davis 2005, 199). However, "a kind of Platonism" already suggests that this is not conclusive evidence.

(Ibid., 51). Gödel then describes the "the lowest of these layers" (Ibid.) in roughly the following terms, which, surprisingly, he doesn't designate as "finitary" or "finitistic", but only by the name "A":

A1: Universal quantification is restricted to infinite totalities for which we can give a finite procedure for generating all their elements (for example, the totality of integers, but not the totality of properties of integers); the main reason for this is that "totalities whose elements cannot be generated by a well-defined procedure are in some sense vague and indefinite as to their borders." (Ibid, p. 53)

A2: Existential statements (and negation of universal ones) are used only as abbreviations, indicating that a particular example has been found without explicitly indicating it; this condition is to be in accord with the constructive notion of existence.

A3: Only decidable notions and calculable functions can be introduced. Since those notion and proof can always be defined by complete induction, system A can be said to be based exclusively on the method of complete induction.

The main point of interest of system A, besides its being the lowest layer of constructive systems, lies in the fact that "this method possesses a particularly high degree of evidence, and therefore it would be the most desirable thing if the freedom from contradiction of ordinary non-constructive mathematics could be proved by methods allowable in this system A. And as

a matter of fact, all the attempts for a proof of freedom from contradiction undertaken by Hilbert and his disciples tried to accomplish exactly that. …. Now all the intuitionistic proofs complying with the requirements of the system A which have ever been constructed can easily be expressed in the system of classical analysis and even in the system of classical arithmetic, and there are reasons for believing that this will hold for any proof which one will ever be able to construct". (Ibid. p.52)

This is the first evidence that seems to suggest that Gödel has come to the same conclusion as von Neumann, with whom he disagreed only two years earlier, that finitary proofs in the sense of Hilbert can be expressed in a single formal system. The change of attitude is surprising, but maybe not as unexpected as it might look. Earlier that year Gödel has found a relative consistency proof of PA over intuitionistic arithmetic HA with his double negation translation. This fact has made it evident that "*the system of intuitionistic arithmetic and number theory is only apparently narrower than the classical one, and in truth contains it, albeit with a somewhat deviant interpretation*" (Gödel, 1933a, 295) At the end of the relative consistency proof paper Gödel also mentioned it explicitly that this proof isn't finitary in the sense of Hilbert, for the reason that "the intuitionistic prohibition against negated universal propositions as purely existential propositions ceased to have any effect because the predicate of absurdity can be applied to universal propositions, and this leads to propositions that formally are exactly the same as those asserted in classical mathematics" (Ibid.,295). That is to say, it violates condition A2 above, since with the introduction of the notion of absurdity the formula $\neg\forall xP(x)$ will have an independent non-constructive interpretation different from the

constructive interpretation of $\exists x \neg P(x)$. In the 1933 lecture, Gödel also discussed the consistency proof of PA in terms of intuitionistic arithmetic as an extension of the system A along constructive lines. However, Gödel is very critical about the value of this proof, not only because it violates condition A2, but mainly A1 in the sense that it involves the very doubtful and vague notion of "all possible proofs" in the intuitive sense which cannot be generated by a finite rule.[233] In the end, though, Gödel expressed hope that "in the future one may find other and more satisfactory methods of construction" (Gödel 1933c, 53) beyond system A to establish the consistency of classical arithmetic and analysis upon them.

The significance of the consistency of PA relative to HA was also recognized quickly by people in the Hilbert school and it convinced Bernays that on the one hand, proof theory, even without sticking to the finitary restrictions, can still be fruitfully developed, and on the other hand, the proof also shows that "the '*finite Standpunkt*' is not the only alternative to classical ways of reasoning and is not necessarily implied by the idea of proof theory. An enlarging of the methods of proof theory was therefore suggested: instead of a restriction to finitist methods of reasoning, it was required only that the arguments be of a constructive character, allowing us to deal with more general forms of inference" (Bernays 1967, 502).

Hilbert himself, however, seems to hold a different opinion about the limits of finitary methods. In his sole reference to Gödel in all of his published works, he writes in the preface to Volume I of *Grundlagen der Mathematik*, published in 1934:

[233] We will discuss in more detail in the next section the value of the consistency proof of PA by HA when we come to compare the three different consistency proofs in terms of their intuitiveness.

> This situation of the results that have been achieved thus far in proof theory at the same time pints the direction for the further research with the end goal to establish as consistent all our usual methods of mathematics.
>
> Regarding this goal, I would like to emphasize that an opinion, which had emerged intermittently—namely that some more recent results of Gödel would imply the infeasibility of my proof theory—has turned out to be erroneous. Indeed, that result shows only that – for more advanced consistency proofs – the finitistic standpoint has to be exploited in a manner that is sharper than the one required for the treatment of the elementary formalism. (Hilbert and Bernays 1934)

Indeed Hilbert does have support on his side. Only two years later in 1936 Gerhard Gentzen proved the consistency of arithmetic by transfinite induction up to the ordinal ε_0 along with other finitary methods. In his paper Gentzen argues for the finitary nature of his whole proof and believes that the proof methods can be considered to be indisputable and thus that "the consistency proof represents a real vindication of the disputable parts of elementary number theory" (Gentzen 1936, 197).

4.3.3 The Zilsel Lecture in 1938

In a remarkable though incomplete seminar presentation in 1938 at the request of Edgar Zilsel [see (Gödel 1938)], Gödel discussed the problem of consistency proof at that time and

suggested three different ways of possible modified extension. He begins by noting first that we can only prove the consistency of partial systems of mathematics as represented in certain particular formal system T, due to the incomplete nature of every modest formal system discovered by himself. However, the existence of comprehensive enough systems like type theory or set theory makes the question of consistency of formal system no less important. Furthermore, every consistency proof is itself a mathematical proof and must be conducted also in a certain system T. However, the question also has an epistemological side. For "after all we want a consistency proof for the purpose of a better foundation of mathematics (laying the foundations more securely)" (Ibid., 89). There are mathematically interesting yet foundationally not so important consistency proofs such as the consistency of PA in second order arithmetic or ZFC by way of a truth definition. For Gödel, "A proof is only satisfying if it either

A. reduces to a proper part or

B. reduces to something which, while not a part, is more evident, reliable, etc., so that one's conviction is thereby strengthened." (Ibid.)

Though (A) is more desirable due to its objective nature by showing superfluous assumptions directly, (B) is no less interesting due to general agreement of the superiority of constructive over non-constructive systems and is actually the route taken historically. But due to the "haziness" of the concept "constructive", Gödel proposed to give a "framework definition, which at least gives necessary if not sufficient conditions". (Ibid., p.91) These conditions are more or less the same as Gödel gave in his 1933 lecture (A1-3), i.e., (1)

functions and relations must be computable and decidable respectively; (2) existence quantifier not as a primitive sign and propositional connectives not to be applied to universal sentences; (3) only laws of the propositional calculus and substitution and ordinary induction be used as rules of inference, and lastly (4) objects should be surveyable (that is, denumerable). Unlike 1933, however, Gödel referred this time to the lowest level of this hierarchy of constructive systems specifically as "finitary number theory", and said that "I believe that Hilbert wanted to carry out the proof of consistency with this" (Ibid. p.93)[234]. And the answer to the question "how far do we get, or fail to get, with finitary number theory" is that transfinite arithmetic (PA)[235] is no longer a part of finitary number theory, for otherwise we could prove the consistency of PA in itself, which is a contradiction (since we can prove the consistency of this finitary number theory (PRA) in PA).

Gödel then proposed three different ways of extension (by violating some of the 4 conditions in one way or another)[236]: the introduction of higher types of functions (functions of functions of numbers, etc.); the modal-logical route and finally transfinite induction. About the first method, which for Gödel has the highest degree of evidence among the three and which will become his main occupation later, in his lecture however, it is only described in a very sketchy way and there is no indication of a concrete method, nor of an interpretation of HA in higher types apart from the negative result that with a fixed finite number of higher types consistency cannot be shown. By the "modal-logical" approach Gödel means the more

[234] Most commentators take the system of this "finitary number theory" to be PRA.
[235] Hilbert mentions laws about the quantifier as "transfinite", compared with laws of propositional connectives. We should not mix this transfinite arithmetic with arithmetic with a transfinite induction.
[236] Feferman in his (Feferman, 2008, p.190) says that all three methods of extension jettisoned the fourth condition about denumerability of objects. However, this assertion is literally wrong about the approach of higher types of functions, at least as it was conceived by Gödel in 1938, where he said explicitly that only this approach "satisfies all four requirements". (Gödel, 1938, p.95/97).

constructive interpretation of Heyting's intuitionistic logic by means of a modal-like operator B (for "Beweisbar", i.e., provable in the absolute, intuitive sense)[237]. After laying down several conditions for B to obtain a constructive system be means of B, Gödel's conclusion is that there is no reasonable way to carry it out and this approach "is the worst of the three ways" (Ibid. p.103) and the degree of evidence it possesses, compared with the system whose consistency is to be proved by it, is "not at all" (Ibid. p.113). The last approach of Gentzen by proving consistency via transfinite induction, lacks the direct evidence as can be gained from finitary number theory. We will discuss Gödel's view of it in more detail in light of the point of view of intuition in the next section.

Maybe the most interesting part of this lecture is the general assessment of Gödel for the significance of the consistency proof by means of this extended approach and of HP in general:

> If the original Hilbert program could have been carried out, that would have been without any doubt of enormous epistemological value. The following requirements would both have been satisfied: (A) Mathematics would have been reduced to a very small part of itself (therefore a large number of independent assumptions would have become superfluous). (B) Everything would really have been reduced to a concrete basis, on which everyone must be able to agree. (Ibid. p. 113)

[237] Gödel has already given an interpretation of intuitionistic logic in terms of the operator B, meaning absolute notion of provability in 1933, without, however, laying stress on any constructive requirements. See (Gödel 1933a)

However, the failure of a consistency proof of PA by finitary number theory already shows that (A) is not possible. As for (B), the three different approaches by means of extended finitism do possess some increase of degree of evidence compared with the system to be proved consistent, to a different extent. However, "it seems to me that the epistemological significance, in the sense of a better foundation, is very much diminished by the fact that the different systems are not contained in finitary number theory" (Ibid. p.113).[238]

In the same year 1938, Bernays was also reporting on the "current question of method in Hilbert's proof theory" (Bernays 1938) paralleling to some extent Gödel's own remarks in Zilsel's lecture. Bernays also considers an extension of the original finitary standpoint to be too narrow but is sceptical about taking over all the intuitionistic methods for a consistency proof. He views, however, Gentzen's consistency proof and its method of transfinite induction up to ε_0 to be intuitively justifiable. He gives an interesting remark that the original finitary standpoint was aroused by the fact that the main problem of proof theory, consistency, can already be formulated in this narrow formalism, thus very likely to be solvable in this formalism too. Gödel's first incompleteness theorem however shows in a definite way that there is a huge gap between the ability to express or formulate a problem and the ability to

[238] Gödel didn't forget to mention that the mathematical significance of extended finitism, "is totally unaffected" and in fact "extraordinarily great". He believes that the methods will lead to "very interesting results in foundational research and also outside it". In a sense, the later development of proof theory has confirmed Gödel's prediction here.

prove and solve it.[239] To the question "what, then, is the characterization of the methodological limitation of proof theory, if not the demand for that elementary evidence which distinguishes the finitary standpoint?", his answer is:

> The tendency to limit methods remains basically the same. However, one may not conceive the evidence and security in an utterly absolute fashion, if one wants to keep open the possibility of extending the methodical framework, then we must avoid using the concepts of evidence and security in too absolute a sense. On the other hand, we thereby gain the principal advantage of not being obliged to question as unjustified or doubtful the usual methods of analysis. (Ibid. p.20)

Not to doubt the "usual methods of analysis" such as non-constructive and impredicative definition seems to contradict directly the aim of HP to finitarily justify those problematical methods using only more intuitive, secure ones. In a letter to Bernays in January 1942, Gödel wrote, referring to the above passage, that: "I read your article … with great interest; only what you say … is not comprehensible to me. Wouldn't that be tantamount to giving up the formalist standpoint? I would be very pleased to hear from you once again" (Gödel 2003a, 133). Bernays replies in the next letter, saying that Gödel's astonishment "is very understandable" given his short elaboration in the 1938 article. His idea is that we don't have to posit, in some absolute sense of what counts as evident methods and take others differing

[239] This interesting remark, by the way, seems to be also "the essential point" of how Gödel viewed his own first incompleteness theorem, see his letter to Zermelo in (Gödel 2003b, 429).

from them as dubious or only technically justified, for that sort of opposition is "not at all necessary in order to do justice to the differences, so long as one resolves to distinguish layers and kinds of evidence" (Ibid.,139). We might also be able to obtain, besides elementary evidence of a specifically arithmetical character, an acquired certainty and evidence in analysis by "intellectual experience" (Ibid.). Martin Davis takes the above exchange of views as evidence that Gödel still attached importance to Hilbert's "formalist standpoint" in 1941 (Davis 2005, 198). However I think Gödel's "astonishment", or surprising reaction is mainly due to the fact that the sort of evidence and certainty to which Bernays referred in the letter no doubt goes beyond Hilbert's original view, and poses a problem for what exactly is the upper bound for Hilbert's notion of intuition and finitism.

4.3.4 The Yale Lecture in 1941/Dialectica Interpretation in 1958/72

In 1941 Gödel delivered a lecture entitled "In what sense is intuitionistic logic constructive?" at Yale university, in which he presented in an informal manner an interpretation of HA in a quantifier-free system of functionals of finite type, the method which he took to be of the highest degree of evidence among all the three approaches of extended finitism considered in the 1938 lecture. 17 years later in 1958, at the occasion of Bernays' 70th birthday, Gödel published in the journal *Dialectica* his functional interpretation of intuitionistic number theory[240], based on the same ideas already in the Yale lecture but

[240] Which has become popularly known since as the "*Dialectica* Interpretation".

changing the title to “On a hitherto unutilized extension of the finitary standpoint”, showing a change of focus and attention of the main aim of this interpretation. A few years later an English translation of this paper (originally published only in German) was prepared and Gödel made an extensive improvement and modification of the original by adding a series of footnotes to it. Under the encouragement of Dana Scott and Bernays the revised manuscript was sent to the printer in 1970. However, when the proof sheets were returned, Gödel was again dissatisfied with some of the added notes and decided not to return for publication. Apparently Gödel continued to work on it until at least 1972, the final form of which was published for the first time in volume II of the *Collected Works* (Gödel 1972a). The full story of the event is told by A. S. Troelstra in his introductory note to Gödel (1958/1972). The amazing twists and turns of this paper indicates, on the one hand, Gödel’s typical extreme caution with his publication, but more importantly, as least as far as this particular paper is concerned, his hesitation about some of his views about Hilbert’s finitism and intuition. We will treat the problems and discussions in these three articles as a whole while concentrating on the differences of focal points.

Notwithstanding Gödel’s assertion in a 1968 letter to Bernays that in 1958 he “placed no particular value on the philosophical matters; rather, I was mainly concerned with the mathematical result, whereas now it is the other way around” (Gödel 2003a, 261), the 1941 Yale lecture first and foremost presents the *Dialectica* interpretation as a foundational contribution, namely, the replacement of abstract intuitionistic concepts by more strictly constructive ones. Gödel notes that there are two independent but yet often confused

objections of the intuitionists against classical mathematicians, namely, objections to impredicative definition and the law of excluded middle, respectively. While the latter seems more serious in its consequence, the double negation translation had already made it abundantly clear that this is only apparently so and that intuitionistic logic "turns out to be rather a renaming and reinterpretation than a radical change of classical logic" (Gödel 1941, 190). The often unnoticed non-constructive element of intuitionistic logic lies in the fact that some of their primitive terms lack the complete perspicuity and clarity which should be required by a constructive system. More specifically, the notion of derivation or of proof as an intuitionistic interpretation of "→" must be taken "in its intuitive meaning as something directly given by intuition, without any further explanation being necessary. This notion of an intuitionistically correct proof or constructive proof lacks the desirable precision" (Ibid.). Then Gödel presents three conditions which he deems to be necessary requirements for any method to be "strictly constructive or finitistic[241]" (Ibid. p.191). They are nearly the same as the conditions in his 1933 or 1938 lecture except for the surveyable (denumerable) condition, i.e., relations and functions decidable, respectively calculable; no existential quantifier as primitive term; and no propositional connectives can be applied to universal propositions. The lowest level of finitistic mathematics is what Gödel called "recursive number theory", where only natural numbers are allowed as objects, and all functions and relations must also be definable by ordinary induction or recursion. This is basically the lowest constructive system as discussed his 1938 lecture, i.e., PRA. Again, this system, though simple and perspicuous, is

[241] Gödel gives an additional explanation for the term "finitistic": "I don't know if the name 'finitistic' is very well chosen, but there is certainly a close relationship between these systems and what Hilbert called the 'finite Einstellung'". (Gödel 1941, 191)

much too weak for the purposes needed in the foundation of mathematics in terms of proving consistency of other systems. Gödel considers the first extension of it along the lines of transfinite induction defined by some well-ordering relation R and mentions Gentzen's consistency proof again, pointing out, however, that the weak point of this method is this: "How does one know that R is a well-ordering without using set-theoretical methods of proof?" (Ibid. p.194). This objection to Gentzen is still consistent with his 1938 one that transfinite induction in Gentzen's proof cannot be made directly intuitive. A more evident extension is introduced by Gödel, allowing the higher type of functions (besides the more usual function of natural numbers) in accord with the schemata of explicit definition and recursive definition. More specifically, we can define a new function F (1) in terms of any term or previously defined functions, or (2) by a recursive definition, i.e., $F(0)=T_1$; $F(x+1)=T_2(x, F(x))$, where T_1 and T_2 are terms composed of previously defined functions and the arguments. Gödel only hints at an informal argument that these higher order functions are actually calculable because they are procedures for obtaining numbers/other procedures out of given numbers/given procedures, and "it is contained in the notion of a procedure that it can always be carried through" (Ibid. p. 195). Gödel then gave a detailed description of this system Σ and how a translation of intuitionistic number theory into Σ is possible, and gave further a few applications of a more mathematical interest and suggested that the interpretation may be extended from arithmetic to give a constructive consistency proof of analysis.

In the publication version of this interpretation in 1958/72[242], rather than addressing it as an interpretation of intuitionistic logic, Gödel puts it in a broader context of extension of the

[242] We will treat two papers together as a single one, though the 1972 paper is much richer in content, and for our

finitary point of view, focusing more on the problem of degree of evidence and the exact bounds of finitism, in Hilbert's sense. This can be seen from the opening paragraph of the paper:

> P. Bernays has pointed out on several occasions that, since the consistency of a system cannot be proved using means of proof weaker than those of the system itself, it is necessary to go beyond the framework of finitary mathematics in Hilbert's sense in order to prove the consistency of classical mathematics, or even that of classical number theory. Since finitary mathematics is defined as the mathematics of concrete intuition, this seems to imply that abstract concepts are needed for the proof of the consistency of number theory. (Gödel 1972a, 271–72)

Finitary mathematics, in the sense of Hilbert, has actually two different components, according to Bernays' observation. One is the constructive element, which only allows mathematical objects or facts that can actually be obtained by a construction or a proof; the other element is the more specific finitistic element, which requires that all mathematical objects and facts should be given in concrete mathematical intuition, that is to say, objects must be finite spatial-temporal configurations of elements. It's the second one which must be modified or abandoned in order to get a suitable system for foundational purposes. Gödel acknowledges the fact that "due to the lack of a precise definition of either concrete or abstract

---

purpose, philosophically more informative, as with most reprints of Gödel's other papers.

evidence, there exists, today, no rigorous proof for the insufficiency (even for the consistency proof of number theory) of finitary mathematics" (Ibid. p. 273). However, Gentzen's proof of the consistency of PA using transfinite induction by ε_0 has made the insufficiency of finitary mathematics "abundantly clear", which is a "surprising fact" (Ibid.). Induction by ε_0 can be proved finitarily if the consistency of number theory could, and "the validity of this induction can certainly not be made immediately evident, as is possible for example in the case of $\omega 2$ ". This is the first evidence in Gödel's publication that seems to put an end to Hilbert's consistency program, contradicting his caution in his 1931 paper about the relationship of his second incompleteness theorem and Hilbert's formalistic standpoint, by putting an upper bound (though not the least) to finitary mathematics, in Hilbert's own sense. The real situation, however, as can been seen from his correspondence with Bernays and from the extensive footnotes to his 1972, presents a more complex picture.

In a letter to Bernays in 1967, Gödel wrote:

> My views have hardly changed since then [that is, 1958], except that I am now convinced that ε_0 is a bound on Hilbert's finitism, not merely in practice but in principle, and that it will also be possible to prove that convincingly. Of course, that does not exclude that there could be (in Hilbert's sense) non-finitary proofs that are equally convincing. (Gödel 2003a, 255)

The main issue, for an upper bound of finitary mathematics, seems to depend on the problem whether recursion on ε_0 can be proved finitarily. In 1969 Bernays sent Gödel a letter along with a new proof of recursion on ε_0 which is to be included in the second edition of *Grundlagen der Mathematik,* Volume 2. Gödel initially must have found the proof very interesting and wrote in a draft letter that:

> … you undoubtedly have given the most convincing proof to date of the ordinal-number character of ε_0. … If one reckons choice sequences to be finitary mathematics, your proof is even finitary. …I think every circularity in the definition of the logical constants really can be avoided. On the other hand I now strongly doubt whether what was said about the boundaries of finitism [in the beginning of his 1958 paper] is really right. For it now seems to me, after more careful consideration, that choice sequences are something concretely evident and therefore are finitary in Hilbert's sense, even if Hilbert himself was perhaps of another opinion. (Ibid.,269)

Gödel was so convinced about Bernays' new proof that he intended to add a footnote to the revised version reporting this result, saying also that "Hilbert did not regard choice sequences (or recursive functions of them) as finitary, but this position may be challenged on the basis of Hilbert's own point of view. (Ibid.)

In the published 1972 version, however, Gödel's reference to choice sequences in footnote (c) expressed a totally different opinion: "A closer approximation to Hilbert's finitism can be achieved by using the concept of free choice sequences instead of 'accessibility'"(Gödel 1972a, 272), but this concept is "really an abstract principle about schemes of ramification", for which even "no satisfactory constructive proof is known" (Ibid.). The change of mind is also reflected in that the above draft letter was never sent, what Gödel really replied is the following, more sceptical letter[243]:

> The proof for the ordinal number character of ε_0 that you give … is extraordinarily elegant and simple. At first one also has the impression that it comes closer to finitism than the other proofs. But on closer reflection that seems very doubtful to me. The property of being "well-founded" contains two quantifiers after all, and one of them refers to all number sequences (which probably are to be interpreted as choice sequences). … It seems to me that one would use a nested recursion for that. But nested recursions are not finitary in Hilbert's sense (i.e., not intuitive), although probably every mathematician will find them just as convincing as primitive recursive definitions. Or don't you believe that? … Hilbert, I presume, didn't want to permit choice sequences? To me they seem to be quite concrete, but not to extend finitism in an essential way. (Ibid. p.271)

[243] William Tait spells out Bernays' proof in detail and shows that it is mistaken or at best incomplete without further assumptions, see (Tait 2006, 89–91).

The reservation Gödel expresses in the letter indicates a real problem for delineating the boundary of finitism, in the sense of Hilbert, i.e., irrespective of what Hilbert actually thought and wrote, is there any bounds for finitary mathematics "on the basis of his [Hilbert's] own point of view", i.e., concrete intuition? Gödel was definitely aware of it, when he writes in 1972 that "whether the necessity of abstract concepts for the proof of induction from a certain point on in the series of constructive ordinals is due solely to the impossibility of grasping intuitively the complicated (though only *finitely* complicated) combinational relations involved, or arises for some essential reason, cannot be decided off hand" (Gödel 1972a, 273–74). This raises the problem that some very complicated relations and concepts, being combinatorial and thus concretely intuitive in Hilbert's sense, might outstrip human being's understanding due to practical reasons and thus appear to be not concretely intuitive *for us*. Gödel did consider the problem of what is possibly the limit of "*idealized*" concrete intuition, abstracting from our practical limitation and with some sort of "reflection" built in:

> Another possibility of extending the original finitary viewpoint … consists in considering as finitary any abstract arguments which only reflect (in a combinatorially finitary manner) on the content of finitary formalisms constructed before, and iterate this reflection transfinitely, using only ordinals constructed in previous stages of this process. (Ibid. p.274)

It's not very clear what the "original finitary viewpoint" is, be it PRA or the more extensive one in accord with the practice of the Hilbert school, but it is definitely weaker than PA. Gödel then mentions an attempt by Kreisel to characterize finitist proof in terms of a transfinite sequence of proof predicates for formal systems, under the condition that this process of ascending is autonomous, i.e., for each possible iteration to stage α there must be an earlier stage β such that the iteration to stage α is finitarily justified.[244] Kreisel shows that ε_0 is limit of this process. Kreisel's result has been known since 1960 and Gödel had already reported it in a letter to Bernays in 1961 that "He [Kreisel] now really seems to have shown in a mathematically satisfying way that the first ε-number is the precise limit of what is finitary. I find this result very beautiful, even if it will require a phenomenological substructure in order to be completely satisfying" (Gödel 2003a, 193). In the published 1972 version, Gödel suspends the question whether this kind of extension can still be called "finitism" in the sense of intuitively evident, but only acknowledges that in this extension (compared with the original Hilbert one) the abstract element appears "in an *essentially weaker form*" than any other extension. As for Kreisel's conclusion that ε_0 is the *exact limit* of idealized concrete intuition, he expresses the opinion that "his arguments would have to be elaborated further in order to be fully convincing" (Gödel 1972a, 274 note f).

4.3.5 The 1961 Lecture Note

[244] For a definitive version, see (Kreisel, 1965, pp.168–173, 177-78).

To complete our survey of Gödel's discussion of HP and finitism, we need to mention the last[245] piece of evidence where Gödel puts his thoughts in a purely philosophical perspective, i.e., the 1961 draft entitled "The modern development of the foundations of mathematics in the light of philosophy" (Gödel 1961), intended as a lecture for the American Philosophical Society, but which was actually never delivered. Gödel first gives a general schema of possible philosophical world-views [*Weltanschauungen*] according to what seems to him "the most fruitful principle" (Ibid. p.375) of the "degree and manner of their affinity to or, respectively, turning away from metaphysics (or religion)" (Ibid.). According to this schema a spectrum of philosophical views can be arranged from the left to right: scepticism, materialism, positivism on the left side, and spiritualism, idealism and theology on the right side. The spirit of time [*Zeitgeist*] has, since the Renaissance, gone largely from right to left and it's most obviously seen in the development of physics, where it's no longer seen as a theoretical science describing an objective states of affairs, but only an instrument in predicting results of observation. Mathematics, however, due to its a priori nature, has long withstood the rule of this spirit of time by evolving into even higher abstractions like set theory and into greater clarity in terms of its foundation. Set-theoretic paradox and other criticism of mathematical reasoning has led, however, most mathematicians to deny classical mathematics as a body of truth, but rather, apart from a certain small part of intuitive evidence, to view most of it merely in a hypothetical sense, i.e., drawing conclusions from certain assumptions without any commitment to truth. Gödel then mentions, as a paradigm example, Hilbert's formalism as a "curious hermaphroditic thing" that "sought to do justice both to the spirit of time and to the

[245] If we consider Gödel's 1972 addition of the 1958 *Dialactica* paper as a whole.

nature of mathematics" (Ibid. p.379). However, for the central aim of Hilbert's formalism, namely, completeness and consistency of the formalism, Gödel's own theorem has made it impossible, in the sense that (a) even if we restrict ourselves to the theory of natural numbers, it is impossible to find a system of axioms and formal rules from which every number-theoretical propositions or its negation will always be derivable and (b) for reasonably comprehensive axioms of mathematics, it is impossible to carry out a proof of consistency "merely by reflecting on the concrete combinations of symbols, without introducing more abstract elements" (Ibid. p.383). Interestingly, Gödel describes this theorem as a reaction of the nature of mathematics itself, which is "very recalcitrant in the face of the *Zeitgeist*" (Ibid. p.379). The rest of the lecture draft is a discussion of possible ways to uphold the rightward view of mathematics in contradiction to the spirit of the time, to which we will return in the next section.

4.4 In the End, What is Gödel's View on HP and Finitism, and How does it Cohere with his Platonism?

Based on the evidence presented in the last section, I will give my own view of Gödel's assessment of HP in particular and finitism in general, with a critique of Davis' judgment in 4.1 and then explain the coherence of Gödel's engagement with finitism and constructive consistency proofs with his own Platonism.

4.4.1 Gödel's View on HP and Finitism

We will discuss below first the (relatively easier) question of how Gödel thinks about HP and secondly the more difficult question of whether he has a stable view on finitism or not. If yes, which one? If not, then why not?

4.4.1.1 Gödel's Disdain for Hilbert's Program?

Martin Davis, after a much briefer summary of more or less the same evidence as our discussion in section 3 above, concluded that:

> Thus over the years Gödel moved from an initial position of allying himself with Hilbert's program, to holding out hope that his own work had not destroyed it, to realizing with some regret that hope was gone, to ultimately speaking of the project with something like disdain. (Davis 2005, 198)

This, I would argue, is a much exaggerated and misleading judgment on Gödel's side. The main argument Davis gives about the so-called "disdain" from Gödel is the phrase "curious hermaphroditic thing" used to describe Hilbert's formalism in Gödel's 1961 note and which Davis ventures to translate as "strange hybrid" as "closer to Gödel's intention". However, in

that note Gödel actually wrote that "as far as the rightness and wrongness, or respectively, truth and falsity, of these two directions [the leftward and rightward in the division of philosophical worldviews] is concerned, the correct attitude appears to me to be that the truth lies in the middle or consists of a combination of the two conceptions", (Gödel 1961, 381) and he acknowledged Hilbert's attempt to be "just such a combination", but only that "too primitive and tending too strongly in one direction" (Ibid.). Nothing here suggests that HP is ill-conceived and Davis's remark doesn't fit very well with Gödel's 1938 paper, that is, after he was pretty sure that HP was a failure, evaluation of HP to be "without any doubt of enormous epistemological value" had it been carried out successfully. A much more coherent and balanced picture can be seen from Gödel's letter to Constance Reid, inquiring the effect of Gödel's incompleteness theorem on HP. Gödel wrote that:

> What has been proved is only that the *specific epistemological objective* which Hilbert had in mind cannot be obtained. This objective was to prove the consistency of the axioms of classical mathematics on the basis of evidence just as concrete and immediately convincing as elementary arithmetic.
>
> However, viewing the situation from a purely mathematical point of view, consistency proofs on the basis of suitably chosen stronger metamathematical presuppositions (as have been given by Gentzen and others) are just as interesting, and they lead to highly important insights into the proof theoretic structure of mathematics. (Reid 1970, 217–18)

Indeed the generalized Hilbert Program in later proof-theoretic developments has certainly produced important mathematical work and given us a much better understanding about the proof structure of mathematics. Though the philosophical significance of these descriptive results may be thin and not directly clear, at least not to the extent Hilbert expected it to be, they might still serve as materials for future philosophical reflection. As Feferman rightly pointed out:

> In general, the kinds of results presented here [the reductive proof-theoretic results] serve to sharpen what is to be said in favor of, or in opposition to, the various philosophies of mathematics such as finitism, predicativism, constructivism, and set-theoretical realism. Whether or not one takes one or another of these philosophies seriously for ontological and/or epistemological reasons, it is important to know which parts of mathematics are in the end justifiable on the basis of the respective philosophies and which are not. (Solomon Feferman 1993, 207)

That is to say, on the one hand, one doesn't have to be a full Platonist to be able to accept most of classical mathematics, as long as they can be reduced to a more constructive part; on the other hand, those non-Platonists can realize more clearly what they cannot get by their principles and be prepared to make those sacrifices. This is well in accord with Gödel's remark

before that certain epistemological positions, in so far as they can be made precise, can be negative refuted or justified, by arguments of mathematical rigour.

4.4.1.2 A Limit for Finitism?

We now come to the more subtle problem of Gödel's view of finitism: whether through the years discussed earlier in section 3 he had a consistent view about finitism or whether he oscillated back and forth between different accounts of it. Thus, for example Feferman in his introductory note to Gödel's correspondence with Bernays speaks about "Gödel's unsettled views over the years as to the exact upper bound of finitary reasoning" (Solomon Feferman 2003, 55), and this has been challenged by Tait, who argues that "Gödel's writings represent a smooth evolution, with just one rather small double-reversal of his view of finitism" (Tait 2010, 88). Before we can give a more definitive answer for this question, a four-fold distinction has to be made about the problem of finitism between:

(a) How should the term "finitism" be understood/used?

(b) How does Hilbert understand/use the term?

(c) How does Gödel take Hilbert to understand/use the term?

(d) How does Gödel understand/use the term?

Among all these four, (a) looks to be a purely philosophical problem while (b) and (d)

appears more to be a historical problem. As it stands, (c), whose answer is essentially for our original question, seems to be partly historical and partly psychological, and thus prone to be of a more speculative nature. However, because of the Hilbertian origin and Gödel's involvement in the exposition of this term as a philosophically important one, the correct answer to (a) is almost certainly bound to be dependent upon answers to (b), (c) and (d).

First of all, what one can say for certain is that both Gödel and Hilbert understand finitism and the related finitary mathematics in terms of some kind of intuitive evidence and certainty, the basic idea of which can again be traced back to Kant. This aspect alone has made Tait's famous analysis of finitism (Tait 1981) seem not that convincing in which he argues that the idea of "iteration" rather than "intuition" is the real basis for finitism, to which PRA is the correct formal system to correspond. Meanwhile, this change of focus in understanding finitism will also affect the epistemological role that it is supposed to play. Rather than being absolutely intuitive and certain, it is only indispensable and irreplaceable:

> . . . no absolute conception of security is realized by finitism or any other kind of mathematical reasoning. Rather, the special role of finitism consists in the circumstance that it is a minimal kind of reasoning presupposed by all nontrivial mathematical reasoning about numbers. And for this reason it is indubitable in a Cartesian sense that there is no preferred or even equally preferable ground on which to stand and criticize it. Thus finitism is fundamental to number-theoretical

mathematics even if it is not a foundation in the sense Hilbert wished. (Tait, 1981, p. 525)

In a late paper about intuition and finitism, Tait indeed tries to establish a link between finitism in his sense (thus PRA) and Kant's philosophy of arithmetic, and goes as far as saying that "on the most plausible reading, a development of Kant's philosophy of arithmetic leads precisely to PRA" (Tait 2010, 95). This seems to bring his argument closer to Hilbert's or Gödel's idea of finitism. However the main argument doesn't go that way: instead of basing the idea of finite iteration on Kantian intuition, he interprets Kantian inner intuition and his concept of magnitude, "on the most plausible reading" to be "finite iteration". So Tait's analysis of finitism as PRA, no matter how plausible it is as a philosophical position about finitism, due to its remote relation to the notion of intuition, doesn't help us much in understanding either Hilbert[246] or Gödel.

Secondly, in Gödel's 1933 and 1938 lecture he mentioned explicitly what he took Hilbert's idea of finitism to be. In the former he said all the methods used in attempts to prove consistency "by Hilbert and his disciples" are all expressible in a system A and in the latter he mentioned that "I believed that Hilbert wanted to carry out the proof with this [finitary number theory]". Now in both cases the systems[247] he referred to can be quite plausibly interpreted to be PRA. Literally speaking Gödel is wrong about this point. What Hilbert himself meant by

[246] To be fair, nor does Tait intend this analysis of finitism as an interpretation of what Hilbert takes finitism to be. If, however the basic ideas and the epistemological role of these two conceptions are so different, then we could wonder why Tait should call his analysis as an analysis of "finitism" rather than something else.

[247] I agree on this point with Tait that the reason Gödel wasn't absolutely sure about the bounds of system A is because of the intuitive notion "finite procedure", which he can already equate with recursively enumerable quite confidently in 1938, after Turing's analysis of mechanical procedure in 1936, and thus the 1933 system A should be PRA too. See (Tait, 2010, p.103)

finitism in theory and in the mathematical practice in the Hilbert school is not totally unambiguous and has lead to more extensive discussions, understandably, than the problem what finitism should be. The trouble is that Hilbert never gave a specific definition to show its limit but rather satisfied himself with informal examples and elucidations[248]. However, ample evidence has been shown by Zach (Zach 1998, 2001) that Hilbert's conception of finitism, at least in so far as its application is concerned, i.e., mathematical practice in pursuit of consistency proofs, definitely goes beyond PRA to include certain functions that are not primitive recursive (such as Ackermann's function or function defined by nested recursion). In the published *Dialectica* paper, however, Gödel was definitely more cautious in delineating what Hilbert's finitism could be. He first noted that we don't have "a precise definition of either concrete or abstract evidence" and cannot have any "rigorous proof for the insufficiency (even for the consistency proof of number theory) of finitary mathematics [in Hilbert's sense]" (Gödel 1972a, 273). Moreover, in the correspondence with Bernays during the 60s and early 70s he showed an ambivalence about what Hilbert took his finitism to be and what Hilbert should take it to be, as can been seen from the discussion respectively about choice sequence, which Gödel took to be concretely intuitive but doubted whether Hilbert will take it so, and nested recursion, which he didn't take to be concretely intuitive but was not very sure about Hilbert's view. Whichever way it goes, it does show that Gödel at that time seemed to attribute something broader than PRA as he did in 1933/38 to Hilbert's finitism. Also true is the fact that in the final 1972 version, where Gödel's discussion of finitism as based on intuition is

[248] Quite understandable for Hilbert qua mathematician. If, as he believed, we could give a finitary consistency proof of analysis, then an actual proof will do and people will recognize its finitary nature. It's only the problem that maybe there does NOT exist such a finitary proof that requires a precise definition. A similar case could be said about the notion of "algorithm".

most elaborate, free choice sequences are regarded as an abstract concept and concrete intuition, even if idealized, still cannot go beyond ε_0, which is the ordinal strength of PA. In a nutshell, Gödel's conception of Hilbertian finitism as mathematics whose evidence rests on concrete intuition never changed through his discussion of finitism; what does change is the exact limit of what this concrete intuition amounts to, i.e., its bound. ε_0 (thus formalizability in PA), for Gödel, is always a bound for this concrete intuition, whether it's the exact least upper bound, he may have oscillated.

The last question about what Gödel himself takes finitism to be is no less complicated, but I think it's fairly safe to say that it goes far beyond Hilbert's notion of finitism, although, unlike in the latter case, we may not be able to set an upper bound for it. Tait, after rejecting attributing to Gödel an unreasonable fluctuation in his views about finitism, suggests that Gödel is unsettled only to the extent "of his view of *Hilbert's* finitism, and the instability centers around his view of whether or not there is or could be a precise analysis of what is 'intuitive'" (Tait 2006, 94). The unsettlement involves on the one hand the contrast of his earlier view in 1931 about the open-endless of finitism to escape formalization in any particular formal system with his later view about the possibility of putting an upper bound, and on the other hand what is exactly the least upper bound for finitism in Hilbert's sense, if there is an upper bound for it. He also argues that Gödel's own view of finitary is stable and can be represented by what he says about the system of finitary number theory in his 1938 Zilsel lecture, which is interpreted to be PRA. I disagree with both of Tait's assertions, to differing extents. For the first, as our discussion above also shows, Gödel indeed wasn't sure about the exact limit of Hilbert's finitism. Initially he took it to be PRA, but believed that it

could be broader to allow more methods of proof. In 1972, although no rigorous proof could be given, Gödel was already describing the insufficiency of finitary consistency proof, in Hilbert's sense, for number theory as a "surprising fact". The same could be said about ε_0 as an upper limit: though no *fully* convincing proof exists, there is abundant evidence for it and Gödel wasn't unsettled either. What I disagree with Tait about Gödel's notion of Hilbert's finitism is that I think the view expressed by Gödel in the 1931 paper about the open-endless nature of a finitary proof should be read as what Gödel himself takes finitism to be, rather than what he takes Hilbert's finitism to be. First of all, this was written in a time before the narrow notion of finitism is distinguished from the more complex notion of intuitionistic proof, whose hallmark is its constructive nature. It's quite natural to take the notion of "constructive" as open-endless and this is actually what Gödel did when he describes Hilbert's finitism as the lowest level of constructive systems. It's only after this distinction that we can separate the two independent elements of finitary mathematics as was pointed out by Bernays, i.e., the general constructive element and the more specific finitistic element. The first is more general in the sense that the notion of a construction, or a proof, in its intuitive sense is unbounded while the second condition is more restrictive in that it only allows objects and facts given in concrete mathematical intuition. And it is exactly the second restriction that Gödel takes issues with, as he expresses clearly in a footnote added for the 1972 paper:

> "Concrete intuition", "concretely intuitive" are used as translations of "Anschauung", "anschaulich"… What Hilbert means by "Anschauung" is substantially Kant's space-time intuition confined, however, to configurations of a

> finite number of discrete objects. Note that it is Hilbert's insistence on concrete knowledge that makes finitary mathematics so surprisingly weak and excludes many things that are just as incontrovertibly evident to everybody as finitary number theory. E.g., while any primitive recursive definition is finitary, the general principle of primitive recursive definition is not a finitary proposition, because it contains the abstract concept of function. There is nothing in the term "finitary" which would suggest a restriction to concrete knowledge. Only Hilbert's special interpretation of it introduces this restriction. (Gödel 1972a, 272)

This, I think, should have made it abundantly clear that Gödel's notion of finitism as based on intuition, both concrete and abstract[249], although admits no rigorous definition, goes much beyond Hilbert's finitism in an essential way. Tait, in interpreting Gödel's finitism as PRA, offers an argument that denies any possibility of such abstract elements in finitism. He first notes that (a) "but surely there is something in the term 'finitary' that suggests a restriction to finite objects" (Tait 2006, 97), and then takes it that (b) "Gödel's distinction between concrete objects and abstract objects coincides extensionally with our distinction between finite objects and infinite objects" (Ibid.). Thus, (a) and (b) together rule out the notion of a finitary mathematics based on abstract elements. But Tait's argument lacks force in the sense that both (a) and (b) are far from being uncontroversial. If finitism is to serve the epistemological role of securing the foundation of mathematics, rather than an ontological or semantic claim[250], then

[249] We will discuss the problem of what counts as "abstract" below. For the purpose of the argument here, it suffices that such things, categorically different from the "concrete" exists for Gödel have a role to play in finitary mathematics.
[250] See (Giaquinto 2002, 156) for a distinction and discussion for the three different senses of finitism.

it's not so obvious why finitary mathematics has to deal only with finite objects but exclude other objects or facts which are as intuitive and evident. Secondly, Gödel never claims that the objects of abstract intuition are infinite objects, but rather "concepts, which are essentially of the second or higher order", and the proof of propositions involving these concepts need insights, not from reflection "upon the combinatorial (space-time) properties of the symbols", but rather "from a reflection upon the *meaning* involved". Nothing in the concept of abstractive intuition implies that it has to have an infinite extension. So what we can say at best is that Tait's conclusion about Gödel's own notion of finitism is what he wants Gödel to take it to be, or what he thinks Gödel should take it to be. The much more plausible interpretation of Gödel's own view of finitism, as far as I can see, is that it's mathematics based on constructive and intuitive principles, which goes beyond Hilbert's finitism and no upper bound is conceived.

4.4.2 Gödel's Platonism and his Concern with Finitism and Constructivism

The last and maybe the most intriguing question is, given his own steady Platonism about mathematics since his student days, how does Gödel's serious engagement[251] with finitism and constructive consistency proofs—as can been seen from discussions above—fit with his own philosophy and belief?

[251] Kreisel in his survey article about Gödel's "excursions into intuitionistic logic" (Kreisel 1987) expressed the view that Gödel didn't really take finitism or constructive consistency project seriously. However, I agree with Feferman that based on the evidence this conclusion seems to "tell us more about Kreisel than about Gödel". (Feferman 2008, 199)

4.4.2.1 Three Possible Solutions

The easiest way to solve this *prima facie* incoherence is to doubt the reliability of Gödel's retrospective view that he held a Platonistic position at such an early age. After all, Gödel's concern with consistency proof and finitism started with his logical career around 1930 and the first public pronouncement of his realistic point of view towards mathematics and logic seemed only to come in 1944, in his famous "Russell's Mathematical Logic" (Gödel 1944). Martin Davis, for example, after examining several evidence in Gödel's lecture notes in the 1930s, tries to put question as to what Gödel actually believed before 1944 (Davis 2005).[252] However, irrespective of the credibility of Davis' arguments, they can certainly not explain the ample and more complicated and mature efforts shown in Gödel (1958/72) and the relevant correspondence with Bernays, in which time Gödel's full-fledged Platonism (especially towards set theory) has become well known.

The second, much more interesting way to reconcile Gödel's views is some psychological explanation from Gödel's point of view. Thus Feferman speaks about the "shadow of Hilbert that loomed over Gödel from the beginning to the end of his career" (Solomon Feferman 2008, 179). Takeuti, in his discussion about Gödel, has expressed the view that not only Hilbert's

[252] Charles Parsons also expressed the view that Gödel's early realism "fell short of what he expressed later [since 1944]" (Parsons 1995).

existence and achievement had a tremendous influence over Gödel, but also that Gödel's career aim was molded to exceed Hilbert (Takeuti 2003 Chapter 3). Unlike Takeuti, who considers Gödel to have achieved no substantial work after 1940 compared with his completeness and incompleteness theorem for logic and arithmetic, the consistency of the continuum hypothesis, Feferman regards Gödel's *Dialectica* interpretation as equally important, and even more so, in the sense that Gödel wants to demonstrate in a decisive way that Hilbert's finitism is never adequate enough to achieve the consistency proof Hilbert wants. Just like Hilbert did to Kronecker before, "to do battle with [him] with his own weapon of finiteness", the goal for Gödel, who "never received the recognition from Hilbert that he deserved", is to "do battle with Hilbert with his own weapon of the consistency program" (Solomon Feferman 2008, 200).

I don't doubt the correctness of such an interpretation, yet I cannot attach to it the same significance that Feferman does. Whether or not Gödel would have received more satisfaction, had Hilbert have acknowledged his work and given him due credit, I don't know, but what appears to be more likely and what I am going to argue is that he doesn't pay so much attention to the problem of finitism and consistence proof, just because he feels he has to do battle against Hilbert using Hilbert's own weapon.

A third possible way is to recognize that, even for an unadulterated Platonist like Gödel, a consistency proof will still have a huge epistemological significance if some non-constructive system can be proved consistent by a constructive one, or a non-evident one by a much more

evident system. This is a very understandable reading, and maybe the most plausible one. But, there are still problems surrounding this interpretation, as Feferman has pointed out that (1) it's curious that Gödel's own engagement in the consistency proof never went beyond arithmetic, and (2) even in the area of arithmetic, Gödel could hardly have thought that a constructive proof would make its consistency more evident than what's provided by the intuitive conception (Solomon Feferman 2008, 200–201). The view I am going to argue for is that, rather than treating the pursuit of finitism and constructive proof as something alien to Gödel's Platonism, the study of the limits of finitism and constructive methods could itself well be an argument for Platonism, although only from a negative point of view. Furthermore, apart from the epistemological perspicuity of a consistency proof, what is more important for what we can learn from the different consistency proofs is the indispensability of the notion of an abstract mathematical intuition, which again is so closely related to Gödelian Platonism. Under this perspective, Feferman's two problems can also be solved in a reasonable way. Let me elaborate.

4.4.2.2 Constructive Proof of Analysis?

First of all, it is to be noted that Gödel did consider the possibility of extending his *Dialectica* interpretation to obtain constructive consistency proofs for systems more comprehensive than number theory, including analysis. Thus near the end of his 1958 paper he stated that "It is clear that … one can also construct systems that are much stronger than T, for

example by admitting transfinite types or the sort of inference that Brouwer used in proving the 'fan theorem'" (Gödel 1958, 251). Indeed, Spector's proof of the consistency of analysis (Spector 1962) using the method of what is called "bar recursion" can be seen as just such an extension. However, the proof can hardly be said to be constructive. This is also pointed out by Gödel in the 1972 revision, where he said in a footnote that for the main principle of Spector's method "no satisfactory constructivistic proof is known" (Gödel 1972a, 272 footnote d) and also suggested a more promising extension by "introducing higher-type computable functions of constructive ordinals". (Ibid.) It's hard to insist like Feferman, with all this evidence, that Gödel never conceived any consistency proof beyond arithmetic. What is more likely is that, just as the case of Continuum Hypothesis (CH) in set theory, where he proved the consistency of CH with Zermelo-Fraenkel but failed to prove the consistency of the negation of CH, thus its independence despite his conviction that CH is independent of ZF, Gödel may have tried all sorts of constructive consistency proof of analysis but failed. That this possibility is important for such a proof for Gödel is also reflected in the same letter to Reid in 1970, where he answered Reid's inquiry about the relationship between his incompleteness theorem and Hilbert's finitary project:

> Moreover, the question remains open whether, or to what extent, it is possible, on the basis of the formalistic approach, to prove 'constructively' the consistency of classical mathematics, i.e., to replace its axioms about the abstract entities of an

objective Platonic realm by insights about the given operations of our mind. (Reid 1970, 217–18)

The way Gödel describes the significance of a constructive consistency proof of analysis is most interesting. In 1938 he had already expressed the view that the success of Hilbert's program would reduce mathematics to a concrete basis, on which everyone must be able to agree (Gödel 1938, 113). Constructive reasoning, on the other hand, is not about objects and facts on a concrete basis, but rather "insights about the given operations of our mind". It seems then, for Gödel, that a constructive consistency proof, besides the epistemological gain of clarity and perspicuity, also has an ontological dimension about the necessity of existence of abstract objects. From this point of view, Gödel's engagement with consistency proofs, rather than just a battle against Hilbert and a bonus of epistemological value, more essentially constitutes part and parcel of his Platonistic philosophy about mathematics. The open question in 1970 may still be open now, but we do have more evidence than before about the likelihood of each position. Per Martin-Löf, in a centenary conference about the birth of intuitionism and constructive mathematics (Martin-Löf 2008), expressed the situation in a most vivid way. He first examined all the efforts, after the failure of Hilbert's program, to obtain a constructive consistency proof for classical analysis (i.e., second-order arithmetic) by whatever constructive principles that can be conceived. He concluded that even though we can never fix absolutely constructive reasoning at any particular level since as soon as you fix it we can always go beyond it by some kind of diagonal or reflection principle, the situation seems to be that we have exhausted all possible constructive methods, and he proposed to call this new situation

"the second failure of the Hilbert program":

> The first failure of the Hilbert program was the one which was discovered by Gödel and of which we are now all aware, but that gave rise to the revised, or modified, Hilbert program, whose characteristic is that we no longer allow merely combinatorial methods in the consistency proof but arbitrarily strong constructive methods. But even this revised, or modified, Hilbert program has come to an end in the nineties, or has failed in the nineties, so it is the second failure of the original Hilbert program, which I cannot interpret in any other way than that we have to give up the dream of being able to establish the consistency of classical mathematics by constructive means. (Martin-Löf 2008, 254)

What Martin-Löf saw from this "second failure" is a triumph of Brouwer over Hilbert, that some non-constructive classical principles like the law of excluded middle could never be justified in a constructive, thus, in Hilbert's point of view, satisfactory way. Viewing this result from a Gödelian point of view, however, it is just a forceful argument for Platonism in that the abstract objects and non-constructive methods postulated in the axioms of classical mathematics (including analysis and set theory) can never be explained away as operations of the mind.

4.4.2.3 Intuition and the Consistency of Number Theory

The second aspect of Gödel's concern with consistency proof, even restricted in the domain of number theory, involves establishing the existence and necessity of what he called "abstraction intuition". It might be true, as Feferman pointed out, that a consistency proof of elementary number theory could never appear more evident than the intuitive notion provided by a Platonistic position. It is nevertheless useful to establish this fact and see what a satisfactory consistency proof could provide and which epistemological attitude will be the most appropriate one in this test case. The general conclusion to be draw from different kinds of consistency proofs for number theory is that abstract intuition and concepts, besides the concrete intuition and methods therewith, is an indispensable part. In the case of the consistency proof by Gentzen, using transfinite induction up to ε_0, it is not satisfactory in Hilbert's sense in that

> the validity of inference by recursion up to ε_0 can certainly not be made *immediately* evident, as it is possible for example in the case of ω^2. That is to say, one cannot grasp at one glance the various structural possibilities which exists for decreasing sequences, and there exists, therefore, no *immediately* concrete knowledge of the termination of every such sequence. But furthermore such concrete knowledge (in Hilbert's sense) cannot be realized either by stepwise transition from smaller to larger ordinal numbers, because the concretely evident

steps, such as $\alpha \to \alpha^2$ are so small that they would have to be repeated ε_0 times in order to reach ε_0. (Gödel 1972a, 273)

What's essential for Gentzen's proof is that concrete intuition is never enough to justify its method, rather we need an abstract notion of "accessibility" or "free choice sequence". As for the consistency proofs of PA in terms of HA, Gödel's main objection is nearly always the same, that the notion of "proof" in intuitionistic logic doesn't have the constructive character and intuitive evidence it is supposed to have. Early in his 1933 lecture he made this point by noting that "Heyting's axioms … differ from the system A [the lowest of constructive mathematics] only by the fact that the substrate on which the constructions are carried out are proofs instead of numbers or other enumerable sets of mathematical objects" (Gödel 1933c, 53), but this very fact itself violates the principle that we should only deal with totalities whose elements can be generated by a finite procedure, which is not the case with the totality of all possible proofs. In fact, this lack of the desirable precision for an intuitionistically correct proof or constructive proof, could well serve as a reduction proof for itself, in the sense that "it furnishes itself a counterexample against its own admissibility, insofar as it is doubtful whether a proof utilizing this notion of a constructive proof is constructive or not" (Gödel 1941, 190). Gödel's philosophical interest in the *Dialectica* interpretation is exactly to replace such abstract yet non-intuitive notions by more evident, constructive ones, which can become the appropriate objection of our abstract intuition. In footnote h of the 1972 version, after a painstaking explanation of the independence of the concept "computable functions of finite type" over the concept "intuitionistic logic" or the concept of proof used by Heyting, Gödel

describes the meaning of this fact as that

> [It] shows that the interpretation of intuitionistic logic, in terms of computable functions, in no way presupposes Heyting's and that moreover, it is constructive and evident in a higher degree than Heyting's. For it is exactly the elimination of such vast generalities as 'any proof' which makes for greater evidence and constructivity. (Gödel 1972a, 276 footnote h)

Apart from arguing for the superiority of the concept of computable functions of finite type over intuitionistic proof in this way, another big change related to this concept from the 1941 Yale lecture to the publication of it in 1958/72 is that rather than define this concept using the two schemes of explicit and recursive definition and then prove that all functions defined in this way are really computable, in the publication version Gödel takes it as a primitive notion directly. Part of the reason is that, as Troelstra mentioned in his introduction to Gödel (1941), Gödel himself realized the circularity problem in the proof: "I don't want to give this proof in more detail because it is of no value for our purpose for the following reason: if you analyze this proof it turns out that it makes use of logical axioms, also for expressions containing quantifiers and … it is exactly these axioms which we want to deduce from the system Σ" (Gödel 1941, 188). As to the possible doubt that whether we have a sufficiently clear idea of the content of the concept computable functions of finite type, Gödel used an example from Turing to argue that the central phrase "well-defined mathematical procedure" in the

explanation of computable functions of finite type can be accepted as having a clear meaning without any further explanation. The example from Turing, is of course his famous analysis of the concept “mechanical computability”:

> As is well known, A.M. Turing has given an elaborate definition of the concept of a *mechanically* computable function of natural numbers. This definition most certainly was not superfluous. However, if the term “mechanically computable” had not had a clear, although unanalyzed, meaning before, the question as to whether Turing’s definition is adequate would be meaningless, while it undoubtedly has an affirmative answer. (Gödel 1972a, 275)

This brings along a central point in Gödel’s notion of abstract intuition, that it is closely related to his own conceptual realism, i.e., abstract concepts are perceived sharply in abstract mathematical intuition. He had already expressed a similar view in his conversation with Hao Wang. As for the question as to how we can find a sharp concept to correspond faithfully to the intuitive one, Gödel’s answer is that “the sharp concept is there all along, only we did not perceive it clearly at first. This is similar to our perception of an animal first far away and then nearby” (Wang 1974, 85), just like we all didn’t perceive the sharp concept of mechanical procedure until Turing brought us to the right perspective. Logically equivalent yet different characterizations of the same concept is the analogue of perceiving a sensory object from different perspectives. What is more, “if there is nothing sharp to begin with, it is hard to

understand how, in many cases, a vague concept can uniquely determine a sharp one without even the slightest freedom of choice". (Ibid.) Obviously he saw the concept of computable functions of finite type as a candidate of such abstract concepts with a definite meaning. By "abstract" he refers to

> …concepts which are essentially of the second or higher level, i.e., which do not have as their content properties or relations of *concrete objects* (such as combinations of symbols), but rather of *thought structures* or *thought contents* (e.g., proofs, meaningful propositions and so on), where in the proofs of propositions about these mental objects insights are needed which are not derived from a reflection upon the combinatorial (space-time) properties of the symbols representing them, but rather upon a reflection upon the *meanings* involved. (Gödel 1972a, 272–73)

For Gödel then, abstract intuition, by reflecting on the meaning of those abstract concepts themselves, will provide us a tool for accomplishing more and more foundational aims in mathematics, such as the consistency proof of formal mathematical systems. And much more can be achieved as well. In a famous passage in his exposition of Cantor's continuum problem, he writes:

> Despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true. I don't see any reason why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense perception, which induces us to build up physical theories and to expect that future sense perceptions will agree with them and, moreover, to believe that a question not decidable now has meaning and may be decided in the future…. That new mathematical intuitions leading to a decision of such problems as Cantor's continuum hypothesis are perfectly possible was pointed out earlier. (Gödel 1964, 268)

The intuition which appears in the above passage seems to be a particular set-theoretic intuition rather than general abstract intuition, and Gödel is talking about "something like a perception also of the objects of set theory", not abstract concepts as in the *Dialectica* interpretation. However, the central ideas are very close. Another occasion where Gödel talks about this intuition, be it abstract or set-theoretical, in a similar but more detailed manner was the 1961 draft lecture which we have mentioned above. After pointing out Hilbert's attempt to be "too primitive and tending too strong in one direction" (Gödel 1961, 381), he also indicates a more promising way of proceeding:

> Obviously, this means that the certainty of mathematics is to be secured not by proving properties by a projection onto material systems—namely, the manipulation

> of physical symbols—but rather by cultivating (deepening) knowledge of the abstract concept themselves which lead to the setting up of these mechanical systems, and further by seeking, according to the same procedures, to gain insights into the solvability, and the actual methods for the solution, of all meaningful mathematical problems. (Ibid., 383)

This act of "cultivating knowledge of abstract concepts" cannot consist of giving explicit definitions on pain of regress and circularity, but "focusing more sharply on the concepts by directing our attention in a certain way, namely, onto our own acts in the use of these concepts, onto our powers in carrying out our acts, etc" (Ibid.). Gödel attributes this method to the phenomenology of Husserl. This explicit reference to Husserl, lacking in any of his published papers, reflects on the one hand his interest and belief in the correctness of Husserl's philosophy[253] for explicating abstract concepts, and on the other hand, the unsatisfactoriness of the extent to which the method has been put in useful application. Another more interesting remark concerns the relation between Husserl and Kant, where Gödel says that "what Husserl did was merely that he first formulated it [central ideas of Kant] more precisely, made it fully conscious and actually carried it out for particular domains" (Ibid., 385). The central idea of Kant, in the case of mathematics, is always the emphasis on the need of intuition, which is what separates the human mathematical mind from a machine. Gödel expressed it as follows:

[253] Gödel started his serious study on Husserl around 1959, see (Wang 1981, 658).

> I would like to point out that this intuitive grasping of ever newer axioms that are logically independent from the earlier ones, which is necessary for the solvability of all problems even within a very limited domain, agrees in principle with the Kantian conception of mathematics. The relevant utterances by Kant, are, it is true, incorrect if taken literally, since Kant asserts that in the derivation of geometrical theorems we always need new geometrical intuitions, that that therefore a purely logical derivation from a finite number of axioms is impossible. That is demonstrably false. However, if in this proposition we replace the term "geometrical" by "mathematical" or "set-theoretical", then it becomes a demonstrably true proposition. I believe it to be a general feature of many of Kant's assertions that literally understood they are false but in a broader sense contain deep truths. (Ibid.)

This is such a striking passage that we need to read more carefully. Tait, in his discussion about Gödel's notion of intuition, remarked that the notion of abstract intuition, which Gödel was "trying to work out—but never succeeded—in the *Dialectica* paper and its revision" (Tait 2010, 94) is a different conception of intuition from the kind of intuition in either the 1961 or 1964 paper, and both are different from the concrete intuition in Hilbert's finitism. The main argument from Tait, quite weak as far as I can see, is that either the intuition in 1961 or the set-theoretical intuition in his 1964 paper is used for grasping and obtaining new axioms, while the abstract intuition in the *Dialectica* paper is mainly supposed to play an important foundational role like consistency proofs and must rest on a "different, non-axiomatic foundation from the

axiomatic theories whose consistency is to be proved" (Ibid.). The main weakness in Tait's argument is that he presupposed a formalistic position, i.e., that the intuition for obtaining a new principle or axioms in some metamathematics for proving the consistency of a formal mathematical system cannot at the same time be the source of new axioms in the formal part as well. The truth and justification for one is its intuitive certainty while the justification or another can only be in a hypothetical sense, that the formal system with a new axioms added is also consistent, guaranteed by the more intuitively certain part. This, however doesn't exclude the possibility of both types of axioms, or at least some of the axioms in the metamathematical part and axioms in the formal part share a common justification, i.e., from the meaning and nature of the mathematical concepts under investigation. The failure of the original HP gives even more support for this reading, in the sense that concrete intuition, void of any consideration of abstract concepts, shall never be adequate as a foundation. That the two concept of intuitions—abstract intuitions for consistency proof and set-theoretic, or more generally mathematical intuition for grasping new axioms to solve undecidable mathematical problems—agree with each other and are essentially the same one can also be seen from Gödel's own remarks. In his suggestion for new axioms to solve Cantor's continuum problem, apart from the usual more powerful axioms of infinity, he also indicated the possibility of finding "other (hitherto unknown) axioms of set theory which a more profound understanding of the concepts underlying logic and mathematics would enable us to recognize as implied by these concepts" (Gödel 1964, 261). It's clear that, despite their different roles in providing a better foundation and deepening the development of mathematics in an essentially way, abstract intuition and set-theoretical intuition have as their content the cultivating and

understanding of abstract concepts, either computable functions of finite type or set for instance, thus on the one hand grounding mathematics by setting up more and more comprehensive and reliable formal mathematical systems and, on the other hand developing it by gaining insights and new methods of proof beyond any particular formal system.

Another issue on which commentators have faulted Gödel is his reference to Kant's view of geometry and intuition in the above quoted passage. Tait, for example concludes that "Gödel's conception of intuition [in 1961 and 1964 paper] has little to do with Kant's notion of intuition" (Tait 2010, 97), and also that Gödel's understanding's of Kant is "a mistake, and it is this incorrect understanding of Kant that supports Gödel's view that reasoning based upon concrete intuition is open-ended"(Ibid., 95). I have argued above in 4.12 that Gödel's view in 1931 about the open-endedness of finitary reasoning is better understood as referring to constructive reasoning or finitary reasoning based on both concrete and abstract intuition, thus Tait's charge against is already undermined. As for Gödel's view of Kant's conception of geometry and intuition, Judson Webb, agreeing with Tait, also said that Gödel's claim was "seriously in error" (Webb 2005, 502). However, even without going far into the problem of what is the correct understanding of Kant[254], for the particular assertion of Kant that the theorem, the sum of the angles of a triangle is two right angles, cannot be derived logically from the axioms but is arrived at "through a chain of inferences guided throughout by intuition", Dagfinn Føllesdal has noticed in his introduction note to Gödel (1961) that Gödel's

[254] Tait claims that "on *the most* plausible reading, a development of Kant's philosophy of arithmetic leads *precisely to* PRA", (Tait 2010, 95, my emphasis), and Webb endeavors to give "a less psychologistic account of Kantian space intuition than Gödel's that respects its strengths while better respecting Kant's text". (Webb 2005, 503) Both these positions, as far as my knowledge goes, doesn't seem to be better than, if not less convincing, Gödel's interpretation.

interpretation of it as implying that the theorem cannot be derived logically from the axioms "is not obviously but is central to some influential modern interpretations of Kant's conception of the role of intuition in mathematics". (Gödel 1961, 367) Besides, Gödel's interpretation of Kant seems to also agree with Hilbert, whose understanding is more important for us. Even if, as Paolo Mancosu has argued for with plenty of evidence quite convincingly, that "there is a striking change of emphasis between the articles from the early 1920s [an empirical (or phenomenistic) conception of finitistic intuition] and the explicit appeal to Kant's pure intuition of the late 1920s" (Mancosu 1998, 145), the mature ideas of finitism and the underlying concept of intuition bear a striking resemblance with the basic idea of Kantian pure intuition, i.e., it constitutes the possibility of theoretical knowledge. This is acknowledged explicitly by Bernays, in one of his paper written in 1928, he writes:

> The 'finitist attitude' demanded by Hilbert as methodological foundation must be characterized epistemologically as some sort of *pure intuition*, because, on the one hand, it is intuitive and on the other hand, it goes beyond what can actually be experienced…. The condition of the finitist attitude present themselves thereby as the conditions *for the possibility of theoretical knowledge of nature*, quite in the sense of the Kantian formulation of the problem.
>
> Once this connection is generally recognized, it will be possible for the basic ideas of the Kantian critique of pure reason to be revived in a new form, detached from its particular historical conditions, from whose bounds theoretical science has freed itself. (Bernays 1928, 7)

In 1931, Hilbert also expressed the view that in the investigation of the principles of mathematics his finitism, representing what he also called "the finite mode of thought", constitutes exactly the *a priori* intuitive element which forms the condition of the possibility of all theoretical knowledge, about whose necessity he finds himself in agreement with Kant:

> Even if today we can no longer agree with Kant in the details, nevertheless the most general and fundamental idea of the Kantian epistemology retains its significance: to ascertain the intuitive *a priori* mode of thought [*Einstellung*], and thereby to investigate the condition of the possibility of all knowledge. (Hilbert 1931b, 1149–50)

The similarity with Gödel's remark that it is "a general feature of many of Kant's assertions that literally understood they are false but in a broader sense contain deep truths" is really striking, although Gödel differs with Hilbert about the scope and nature of this "intuition". In a most interesting remark near the end of his 1961 text, Gödel closes the draft with the sentence: "if the misunderstood Kant has already led to so much that is interesting in philosophy, and also indirectly in science, how much more can we expect it from Kant understood correctly?" (Gödel 1961, 387). Indeed, the philosophical conclusion of the necessity of the existence of abstract concepts and mathematical/abstract intuition evidenced

by logical and mathematical investigations, as can be seen from the failure of consistency proof for mathematical systems by pure concrete intuition based on combinatorial relations and manipulations of symbols on one side, and the success and fruitfulness of the introduction of abstract intuition and concepts on the opposite side, maybe has just presented itself as a small step in the bigger project of understanding Kant correctly.

Conclusion

An extremely neat and concise summary of how Gödel himself saw the philosophical consequences of his logical results for general philosophy and Hilbert's program, in particular, is expressed in a letter from Gödel to Leon Rappaport in 1962:[255]

> Nothing has been changed lately in my results or their philosophical consequences, but perhaps some misconceptions have been dispelled or weakened. My theorems only show that the mechanization of mathematics, i.e. the elimination of the mind and of abstract entities, is impossible, if one wants to have a satisfactory foundation and system of mathematics.
>
> I have not proved that there are mathematical questions undecidable for the human mind, but only that there is no machine (or blind formalism) that can decide all number theoretical questions (even of a very certain special kind).
>
> Likewise it does not follow from my theorems that there are no convincing consistency proofs for the usual mathematical formalisms, notwithstanding that such proofs must use modes of reasoning not contained

[255] This passage was brought to my attention while reading (Mancosu 2004).

> in those formalism. What is practically certain[256] is that there are for the classical formalisms, no conclusive combinatorial consistency proofs (such as Hilbert expected to give), i.e. no consistency proofs that use only concepts referring to finite combinations of symbols and not referring to any infinite totality of such combinations. (Gödel 2003b, 176)

All our discussions in the last four chapters, focusing on the more "local arguments", can just be seen as the warp and woof of the picture sketched in Gödel's letter. Although we have not clarified (and may well never be able to clarify) the concepts and questions sufficiently so as to "conduct these discussions with mathematical rigor" (Gödel 1951, 322) and show "that the Platonistic view is the only one tenable" (ibid.) as Gödel expected, we have certainly gone a long distance in confirming the above view by both rejecting as unfeasible certain competing views and gaining credence from positive mathematical and logical results such as Turing's clarification of the concept of "mechanical procedure" and Gödel's consistency proof for PA using his *Dialectica Interpretation*.

More specifically, we have seen that Carnap's view of mathematics as syntactical conventions and thus void of content cannot be successful defended as a foundational thesis. Even as a proposal, it doesn't reflect the true situation of mathematical and logical practice and is in a sense self-refuting because of the

[256] Gödel added a footnote, writing that "no formal proof has yet been given because the concept of a combinatorial proof, although intuitively clear, has not yet been precisely defined".

existence of other more convenient and fruitful ones. The failure of such an attempt indicates the more plausible view that mathematics does have an irreducible content just like physical theory and that the belief in the correctness of mathematical intuition in either pure or applied mathematics cannot be replaced by symbolic conventions. Quine's similar objections against Carnap's linguistic view of logical truth is strengthened in an essential way by Gödel by dropping the requirement of the infinity of logical truths, as is the same with the indispensability argument for Platonism usually attributed to him and Putnam by contrasting it with Gödel's own version. The difficulties in Carnap's tolerance principle and the consequent trivialization of foundational debates are pointed out, showing that justification, more than merely choice, is a necessary part of foundational discussions. In the case of Russell, it manifests itself clearly that it's exactly Russell's reluctance to taking a more realistic position towards mathematical objects and concepts that brings about the main difficulties in his logical system, and this again indirectly confirms the view that logic and mathematics do have an abstract content that cannot be explained away using only concrete materials. The nature of paradoxes and the possible solutions for them also depend on the view taken and Russell didn't separate the general requirement of a solution and his particular constructivistic view, as is shown clearly in our discussion of the vicious circle principle. The theory of simple types, however, is much more plausible under a certain interpretation. As a foundation for mathematics, it reveals in a plain way the inexhaustible nature of

mathematics by allowing the construction of types to an arbitrarily high level without ever coming to an end; as a theory of concepts conceived as objectively existing entities it can be regarded as a stepping stone for a more satisfactory theory. The case study of computability in the third chapter can be seen as a paradigm for conceptual analysis. Logically equivalent characterizations of the same concept don't have to be (and usually are not) epistemologically equal. Some of them might strike one as not only psychologically more intuitive, but also conceptually more convincing. This possibility shows the robust existence of concepts and also should encourage us to find more successful applications of other epistemologically important concepts such as provability. The precise definition of mechanical procedure also makes possible the characterization of the limits of pure formalism and mechanism. I have argued that rather than conflicting with each other, both Turing and Gödel realized the existence of something in the human mind that surpasses purely mechanical methods and they only differed in the different ways in which they tried to bring this out, i.e., by intuition and cultivation of abstract concepts or the idea of machine-learning. The insufficiency of purely finite combinatorial concepts and the necessity and usefulness of abstract concepts is shown again in the particular cases of consistency proofs. By delimiting finitism and the underlying concrete intuition in a precise way, we can then more rigorously prove the necessity of abstract intuition transcending this concrete one.

With all these considerations I think we have demonstrated the high

plausibility of Gödel's view that there do exist abstract mathematical concepts, whose relations constitute the inexhaustible content of mathematics, which again can be perceived by the human mind through mathematical intuition. They are not conclusive arguments, of course, and we have to rely on further results concerning the foundation of mathematics, especially the fate of CH and also elucidations from Gödel's other unpublished philosophical notebooks [257] for making several points in a more convincing way.

We will conclude with a remark from Emil Post, which expresses some of the central tenets of Gödel's philosophy in a remarkable way:

> The conclusion is unescapable that even for such a fixed, well defined body of mathematical propositions, *mathematical thinking is, and must remain, essentially creative.* To the writer's mind, this conclusion must inevitably result in at least a partial reversal of the entire axiomatic trend of the later nineteenth and early twentieth centuries, with a return to meaning and truth as being of the essence of mathematics. (Post 1944, 316)

[257] For an overview of Gödel's *Nachlass*, see (Dawson 2016; Engelen and Crocco 2016).

Bibliography

Ackermann, Wilhelm. 1928. “Zum Hilbertschen Aufbau der reellen Zahlen.” *Mathematische Annalen* 99 (1): 118–33.

Atten, Mark van, and Juliette Kennedy. 2003. “On the Philosophical Development of Kurt Gödel.” *The Bulletin of Symbolic Logic* 9 (4): 425–76.

Atten, Mark van, and Juliette Kennedy. 2009. “‘Gödel’s Modernism: On Set-Theoretic Incompleteness,’ Revisited.” In *Logicism, Intuitionism, and Formalism*, 303–55. Dordrecht: Springer Netherlands.

Avigad, Jeremy, and Erich H. Reck. 2001. “‘Clarifying the Nature of the Infinite’: The Development of Metamathematics and Proof Theory.” *Carnegie Mellon Technical Report CMU-PHIL-120*, 1–53.

Awodey, Steve, and A. W. Carus. 2003. “Carnap vs. Gödel: On Syntax and Tolerance.” In *Logical Empiricism: Historical and Contemporary Perspectives*, edited by P. Parrini, W. C. Salmon, and M. H. Salmon, 57–64. University of Pittsburgh Press.

———. 2004. “How Carnap Could Have Replied to Gödel?” In *Carnap Brought Home: The View from Jena*, 199–220. Open Court.

———. 2009. “From Wittgenstein’s Prison to the Boundless Ocean: Carnap’s Dream of Logical Syntax.” In *Carnap’s Logical Syntax of Language*, edited by Pierre Wagner, 79–

108. Palgrave Macmillan.

———. 2010. “Gödel and Carnap.” In *Kurt Gödel: Essays for His Centennial*, edited by Solomon. et al. Feferman, 252–74. Cambridge University Press.

Benacerraf, Paul. 1965. “What Numbers Could Not Be.” *The Philosophical Review* 74 (1): 47–73.

Bernays, Paul. 1928. “On Nelson’s Position in the Philosophy of Mathematics.” http://www.phil.cmu.edu/projects/bernays/Pdf/bernays07_2004-02-15.pdf.

———. 1935a. “Hilbert’s Invenstigations of the Foundations of Arithmetic.” http://www.phil.cmu.edu/projects/bernays/Pdf/bernays14_2003-05-08.pdf.

———. 1935b. “On Platonism in Mathematics.” In *Philosophy of Mathematics: Selected Readings*, edited by Hilary Putnam and Paul Benacerraf, 258–71. Cambridge University Press.

———. 1938. “On Current Methodological Questions of Hilbert’s Proof Theory.” http://www.phil.cmu.edu/projects/bernays/Pdf/bernays16_2002-11-26.pdf.

———. 1946. “Review: Russell’s Mathematical Logic by Kurt Gödel.” *The Journal of Symbolic Logic* 11 (3): 75–79.

———. 1961. “On the Role of Language from an Epistemological Point of View.” http://www.phil.cmu.edu/projects/bernays/Pdf/bernays25_2003-10-20.pdf.

———. 1967. “Hilbert, David.” In *Encyclopdia of Philosophy, Vol. 3*, edited by P. Edwards, 496–504. Macmillan, New York.

Beth, Willem Evert. 1963. “Carnap’s Views on the Advantages of Constructed Systems over Natural Languages in the Philosophy of Science.” In *The Philosophy of Rudolf Carnap*, edited by Paul Schilpp, 469–502. Open Court.

Black, Robert. 2000. “Proving Church’s Thesis.” *Philosophia Mathematica* 8 (3): 244–58.

Boolos, George. 1995. “Introductory Note to 1951.” In *Kurt Gödel : Collected Works, Vol. III*, 290–304. Oxford University Press.

Burgess, John. 2013. “Quine’s Philosophy of Logic and Mathematics.” In *A Companion to W.V.O. Quine*, edited by Gibbert Harman and Ernie Lepore, 279–95. John Wiley & Sons, Inc.

———. 2014. “Intuitions of Three Kinds in Gödel’s Views on the Continuum.” In *Interpreting Gödel*, edited by Juliette Kennedy, 11–31. Cambridge University Press.

Burgess, John, and A. P. Hazen. 1998. “Predicative Logic and Formal Arithmetic.” *Notre Dame Journal of Formal Logic* 39 (1): 1–17.

Carnap, Rudolf. 1931. “The Logicist Foundations of Mathematics.” In *Philosophy of Mathematics: Selected Readings*, edited by Paul Benacerraf and Hilary Putnam, 2nd Edi., 41–51. Cambridge University Press.

———. 1935. “Formal and Factual Science.” In *Readings in the Philosophy of Science*, edited

by May Brodbeck and Herbert Feigl, 123–28.

———. 1937. *The Logical Syntax of Language*. London: K. Paul, Trench, Trubner & Co.

———. 1942. *Introduction to Semantics*. Harvard University Press.

———. 1950a. “Empiricism, Semantics, and Ontology.” *Revue Internationale de Philosophie*.

———. 1950b. *Logical Foundations of Probability*. Chicago: University of Chicago Press.

———. 1952. “Meaning Postulates.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 3 (5): 65–73.

———. 1963. “Intellectual Autobiography.” In *The Philosophy of Rudolf Carnap*, edited by Paul Schilpp, 3–84.

Carroll, Lewis. 1895. “What the Tortoise Said to Achilles.” *Mind* 4 (14): 278–80.

Cassou-Noguès, Pierre. 2005. “Gödel and ‘the Objective Existence’ of Mathematical Objects.” *History and Philosophy of Logic* 26 (3): 211–28.

Chihara, Charles. 1982. “A Gödelian Thesis Regarding Mathematical Objects: Do They Exist? And Can We Perceive Them?” *The Philosophical Review* 91 (2): 211–27.

———. 1990. *Constructibility and Mathematical Existence*. Oxford: Clarendon Press.

Church, Alonzo. 1932. “A Set of Postulates for the Foundation of Logic.” *Annals of Mathematics* 33 (2): 346–66.

———. 1933. “A Set of Postulates For the Foundation of Logic.” *The Annals of Mathematics* 34 (4): 839.

———. 1934. “The Richard Paradox.” *The American Mathematical Monthly* 41 (6): 356–61.

———. 1935. “An Unsolvable Problem of Elementary Number Theory (Abstract).” *Bulletin of the American Mathematical Society* 41: 332–33.

———. 1936a. “A Note on the Entscheidungsproblem.” In *The Undecidable*, edited by Martin Davis, 108–14.

———. 1936b. “An Unsolvable Problem of Elementary Number Theory.” In *The Undecidable*, edited by Martin Davis, 88–107. Raven Press.

———. 1937a. “Review: Finite Combinatory Processes-Formulation 1.” *The Journal of Symbolic Logic* 2 (1): 43.

———. 1937b. “Review: On Computable Numbers, with an Application to the Entscheidungsproblem.” *The Journal of Symbolic Logic* 2 (1): 42–43.

———. 1943. “Review of Carnap’s ‘Introduction to Semantics.’” *Philosophical Review* 52: 298–304.

———. 1956. *Introduction to Mathematical Logic: Volume 1*. Princeton: Princeton University Press.

Coffa, Alberto. 1987. “Carnap, Tarski and the Search for Truth.” *Noûs* 21 (4): 547–72.

Cohen, Paul. 1963. “The Independence of the Continuum Hypothesis.” *Proceedings of the National Academy of* 50 (6): 1143–48.

Copeland, Jack. 2002. “The Church-Turing Thesis.” *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/church-turing/.

———. , ed. 2004. *The Essential Turing*. Oxford: Clarendon Press.

———. 2006. “Turing’s Thesis.” In *Church’s Thesis after 70 Years*, 147–74.

Copi, Irving M. 1971. *The Theory of Logical Types*. Routledge and Kegan Paul.

Crocco, Gabriella. 2003. “Gödel, Carnap and the Fregean Heritage.” *Synthese* 137 (1/2): 21–41.

———. 2012. “Gödel, Leibniz and ‘Russell’s Mathematical Logic.’” In *New Essays on Leibniz Reception*, 217–56. Basel: Springer Basel.

Davidson, Donald. 1969. “True to the Facts.” *The Journal of Philosophy* 66 (21): 748–64.

Davis, Martin, ed. 1965. *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*. Raven Press.

———. 1973. “Hilbert’s Tenth Problem Is Unsolvable.” *The American Mathematical Monthly* 80 (3): 233–69.

———. 1978. “What Is a Computation?” In *Mathematics Today: Twelve Informal Essays*, 241–67. Springer.

———. 1982. “Why Gödel Didn’t Have Church’s Thesis.” *Information and Control* 54 (1–2): 3–24.

———. 2005. “What Did Gödel Believe and When Did He Believe It?” *The Bulletin of Symbolic Logic* 11 (2): 194–206.

Dawson, John W. 1984. “The Reception of Gödel’s Incompleteness Theorems.” *Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1984: 253–71.

———. 2006. “Gödel and the Origins of Computer Science.” In *Logical Approaches to Computational Barriers*, 133–36. Springer.

———. 2016. “What Have We Learned From the Gödel Nachlass, and What More May It Have to Offer?” In *Kurt Gödel: Philosopher-Scientist*, 15–32.

Detlefsen, Michael. 1986. *Hilbert’s Program: An Essay on Mathematical Instrumentalism*. Reidel.

———. 1992. “Poincaré against the Logicians.” *Synthese* 90 (3): 349–78.

Engelen, Eva-Maria, and Gabriella Crocco. 2015. *Kurt Gödel: Philosopher-Scientist*. Presses Universitaires de Provence.

———. 2016. “Kurt Gödel’s Philosohical Remarks (Max Phil).” In *Kurt Gödel: Philosopher-Scientist*, 34–54.

Feferman, Solomon. 1960. “Arithmetization of Metamathematics in a General Setting.”

Fundamenta Mathematicae 49 (1): 35–92.

———. 1964. “Systems of Predicative Analysis.” *The Journal of Symbolic Logic* 29 (1): 1–30.

———. 1984. “Kurt Gödel: Conviction and Caution.” In *In the Light of Logic*, 150–64. Oxford University Press.

———. 1988. “Turing in the Land of O(z).” In *The Universal Turing Machine: A Half-Century Survey*, 103–34.

———. 1993. “What Rests on What? The Proof-Theoretic Analysis of Mathematics.” In *In the Light of Logic*, 284–98. Oxford University Press.

———. 1995. “Introductory Note to 1933c.” In *Kurt Gödel : Collected Works, Vol. III*, 36–44.

———. 2003. “Introductory Note.” In *Kurt Gödel : Collected Works. Volume IV, Correspondence A-G*, 41–78.

———. 2008. “Lieber Herr Bernays!, Lieber Herr Gödel! Gödel on Finitism, Constructivity and Hilbert’s Program.” *Dialectica* 62 (2): 179–203.

Fitch, Frederic B. 1938. “The Consistency of the Ramified Principia.” *The Journal of Symbolic Logic* 3 (4): 140–49.

Franzén, Torkel. 2005. *Gödel’s Theorem: An Incomplete Guide to Its Use and Abuse*. A K Peters.

Frege, Gottlob. 1902. “Letter to Russell.” In *From Frege to Gödel: A Source Book in*

Mathematical Logic, edited by Jean van Heijenoort, 126–28. Harvard University Press.

———. 1980. *Philosophical and Mathematical Correspondence*. Edited by Gottfried Gabriel, Hans Hermes, and Christian Thiel. Basil Blackwell.

———. 1982. *Philosophical and Mathematical Correspondence*. Edited by G. Gabriel, B. McGuinness, and H. Kaal. Basil Blackwell.

Friedman, Michael. 1988. "Logical Truth and Analyticity in Carnap's Logical Syntax of Language." In *History and Philosophy of Modern Mathematics*, edited by Philip Kitcher and William Aspray, 82–94.

———. 2001. "Tolerance and Analyticity in Carnap's Philosophy." In *Future Pasts: The Analytic Tradition in Twentieth-Century Tradition*, edited by Juliet Floyd and Sanford Shieh, 223–56. Oxford University Press.

Gandy, Robin. 1980. "Church's Thesis and Principles for Mechanisms." In *Studies in Logic and the Foundations of Mathematics: The Kleene Symposium*, 123–48.

———. 1988. "The Confluence of Ideas in 1936." In *The Universal Turing Machine a Half-Century Survey*, edited by R. Herken, 55–111.

Gentzen, Gerhard. 1933. "On the Relation between Intuitionistic and Classical Arithmetic." In *The Collected Papers of Gerhard Gentzen*, edited by M. Szabo, 53–67.

Gentzen, Gerhard. 1936. "The Consistency of Elementary Number Theory." In *The Collected Papers of Gerhard Gentzen*, edited by M. Szabo, 132–213.

Giaquinto, Marcus. 2002. “The Search for Certainty : A Philosophical Account of Foundations of Mathematics.” Clarendon Press.

Gödel, Kurt. 1929. “On the Completeness of the Calculus of Logic.” In *Gödel (1986), Collected Works I*, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 60–101. Oxford University Press.

———. 1931. “On Formally Undecidable Propositions of Principia Mathematica and Related Systems I.” In *Gödel (1986), Collected Works I*, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 144–95. Oxford University Press.

———. 1933a. “An Interpretation of the Intuitionistic Propositional Calculus.” In *Gödel (1986), Collected Works I*, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 301–2. Oxford University Press.

———. 1933b. “On Intuitionistic Arithmetic and Number Theory.” In *Gödel (1986), Collected Works I*, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 286–96. Oxford University Press.

———. 1933c. “The Present Situation in the Foundations of Mathematics.” In *Gödel (1995), Collected Works, Vol III*, edited by Soloman Feferman, John Dawson, Warren Goldfarb,

Charles; Parsons, and Robert M. Solovay, 45–53. Oxford University Press.

———. 1934. “On Undecidable Propositions of Formal Mathematical Systems.” In *Kurt Gödel : Collected Works, Vol. II*, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 346–71. Oxford University Press.

———. 1936. “On the Length of Proofs.” In *Kurt Gödel : Collected Works, Vol. I*, 397–99. Oxford University Press.

———. 1937. “Undecidable Diophantine Propositions.” In *Kurt Gödel: Collected Works, Vol. III*, edited by Soloman Feferman, John Dawson, Warren Goldfarb, Charles Parsons, and Robert M. Solovay, 164–74. Oxford University Press.

———. 1938. “Lecture at Zilsel’s.” In *Gödel (1995), Collected Works, Vol III*, edited by Soloman Feferman, John Dawson, Warren Goldfarb, Charles Parsons, and Robert M. Solovay, 62–113. Oxford University Press.

———. 1940. “The Consistency of the Axiom of Choice and of the Generalized Continuum Hypothesis with the Axioms of Set Theory.” In *Collected Works, Vol. II*, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 33–101. Oxford University Press.

———. 1941. “In What Sense Is Intuitionistic Logic Constructive?” In *Gödel (1995), Collected Works, Vol III*, edited by Soloman Feferman, John Dawson, Warren Goldfarb, Charles Parsons, and Robert M. Solovay, 186–200. Oxford University Press.

———. 1944. “Russell’s Mathematical Logic.” In *Kurt Gödel : Collected Works, Vol. II*, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 119–43. Oxford University Press.

———. 1946. “Remarks before the Princeton Bicentennial Conference on Problems in Mathematics.” In *Kurt Gödel : Collected Works, Vol. II*, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 150–53. Oxford University Press.

———. 1947. “What Is Cantor’s Continuum Problem?” In *Kurt Gödel: Collected Works, Vol. II*, edited by Soloman; Feferman, John; Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 176–88. Oxford University Press.

———. 1949. “A Remark about the Relationship between Relativity Theory and Idealistic Philosophy.” In *Kurt Gödel : Collected Works, Vol. II*, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 202–7. Oxford University Press.

———. 1951. “Some Basis Theorems on the Foundations of Matheamtics and Their Implications.” In *Gödel (1995), Collected Works, Vol III*, edited by Soloman Feferman, John Dawson, Warren Goldfarb, Charles Parsons, and Robert M. Solovay, 304–23. Oxford University Press.

———. 1953a. “Is Mathematics Syntax of Language? Version III.” In *Kurt Gödel : Collected Works, Vol. III*, edited by Soloman Feferman, John Dawson, Warren Goldfarb, Charles

Parsons, and Robert M. Solovay, 334–55. Oxford University Press.

———. 1953b. “Is Mathematics Syntax of Language? Version V.” In *Kurt Gödel : Collected Works, Vol. III*, edited by Soloman Feferman, John Dawson, Warren Goldfarb, Charles Parsons, and Robert M. Solovay, 356–62. Oxford University Press.

———. 1958. “On a Hitherto Unutilized Extension of the Finitary Standpoint.” In *Kurt Gödel : Collected Works, Vol. II*, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 241–53. Oxford University Press.

———. 1961. “The Modern Development of the Foudnations of Mathematics in the Light of Philosophy.” In *Gödel (1995), Collected Works, Vol III*, edited by Soloman Feferman, John Dawson, Warren Goldfarb, Charles Parsons, and Robert M. Solovay, 363–86. Oxford University Press.

———. 1964. “What Is Cantor’s Continuum Problem?” In *Gödel (1990), Collected Works II*, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 254–70. Oxford University Press.

———. 1972a. “On an Extension of Finitary Mathematics Which Has Not yet Been Used.” In *Gödel (1990), Collected Works, Vol. 2*, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 271–80.

———. 1972b. “Some Remarks on the Undecidability Results.” In *Kurt Gödel. Collected*

Works. Volume II Publications 1938-1974, edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort, 305–7. Oxford University Press.

———. 1986. *Kurt Gödel. Collected Works. Volume I: Publications, 1929-1936.* Edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort. Oxford University Press. doi:10.1086/374125.

———. 1990. *Kurt Gödel. Collected Works. Volume II Publications 1938-1974.* Edited by Soloman Feferman, John Dawson, Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort. Oxford University Press.

———. 1995. *Kurt Gödel: Collected Works: Volume III: Unpublished Essays and Lectures*. Oxford University Press.

———. 2003a. *Kurt Gödel : Collected Works. Volume IV, Correspondence A-G*. Edited by Solomon Feferman, John W. Dawson, Warren Goldfarb, Charles Parsons, and Wilfried Sieg. Oxford University Press.

———. 2003b. *Kurt Gödel : Collected Works. Volume V, Correspondence H-Z.* Edited by Solomon Feferman, John W. Dawson, Warren Goldfarb, Charles Parsons, and Wilfried Sieg. Oxford University Press.

Goldfarb, Warren. 1988. “Poincaré against the Logicists.” In *History and Philosophy of Modern Mathematics*, edited by Philip Kitcher and William Aspray, 61–81.

———. 1995. “Introductory Note to *1953/9.” In *Kurt Gödel : Collected Works, Vol. III*, edited by Soloman Feferman, John Dawson, Warren Goldfarb, Charles Parsons, and Robert M. Solovay, 324–34. Oxford University Press.

———. 2005. “On Gödel’s Way In: The Influence of Rudolf Carnap.” *The Bulletin of Symbolic Logic* 11 (2): 185–93.

Goldfarb, Warren, and Thomas Ricketts. 1992. “Carnap and the Philosophy of Mathematics.” In *Science and Subjectivity*, edited by D. Bell and W. Vossenkuhl, 61–78. Berlin: Akadamie Verlag.

Goodstein, R. L. 1944. “On the Restricted Ordinal Theorem.” *The Journal of Symbolic Logic* 9 (2): 33–41.

Hempel, Carl. 1945. “On the Nature of Mathematical Truth.” *The American Mathematical Monthly* 52 (10): 543–56.

Hilbert, David. 1899. “Grundlagen Der Geometrie.” Leipzig.

———. 1900a. “Mathematical Problems.” In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics, Vol. 2*, edited by William Ewald, 1096–1104. Clarendon Press.

———. 1900b. “On the Concept of Number.” In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics, Vol. 2*, edited by William Ewald, 1089–95. Oxford University Press.

———. 1918. “Axiomatic Thought.” In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics, Vol. 2*, edited by William Ewald, 1105–14. Clarendon Press.

———. 1922. “The New Grounding of Mathematics.” In *From Kant to Hilbert: A Source Book in the Foundations 2*, edited by William Ewald, 1115–34. Oxford University Press.

———. 1923. “The Logical Foundations of Mathematics.” In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics, Vol. 2*, edited by William Ewald, 1134–48.

———. 1925. “On the Infinite.” In *From Frege to Gödel: A Source Book in Mathematical Logic*, edited by Jean van Heijenoort, 367–92. Harvard University Press.

———. 1927. “The Foundation of Mathematics.” In *From Frege to Gödel: A Source Book in Mathematical Logic*, edited by Jean van Heijenoort, 464–79. Harvard University Press.

———. 1931a. “The Grounding of Elementary Number Theory.” In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics, Vol. 2*, 485–94.

———. 1931b. “The Grounding of Elementary Number Theory.” In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics, Vol. 2*, edited by William Ewald, 1148–57.

Hilbert, David, and Wilhelm Ackermann. 1950. *Principles of Mathmatical Logic*. Edited by Robert E. Luce. Chelsea Publishing Company.

Hilbert, David, and Paul Bernays. 1934. *Grundlagen Der Mathematik, Vol. 1*. *Springer-Verlag*.

———. 1939. *Grundlagen Der Mathematik, Vol. 2 (1939)*. Springer-Verlag.

Hodges, Andrew. 1983. *Alan Turing: The Enigma*. Burnett Books with Hutchinson.

———. 1988. “Alan Turing and the Turing Machine.” In *The Universal Turing Machine a Half-Century Survey*, 3–15.

James, E. 1992. “The Problem of Mathematical Existence.” *Philosophical Books* 33 (3): 129–38.

Jeroslow, R. G. 1973. “Redundancies in the Hilbert-Bernays Derivability Conditions for Gödel’s Second Incompleteness Theorem.” *The Journal of Symbolic Logic* 38 (3). Association for Symbolic Logic: 359–67.

Kennedy, Juliette. 2013. “On Formalism Freeness: Implementing Gödel’s 1946 Princeton Bicentennial Lecture.” *The Bulletin of Symbolic Logic* 19 (3): 351–93.

———. 2017a. “Turing, Gödel and the ‘Bright Abyss.’” In *Philosophical Explorations of the Legacy of Alan Turing*, 63–91. Springer.

———. 2017b. “Gödel’s Reception of Turing’s Model of Computability: The ‘Shift of Perception’ in 1934.” In *Unveiling Dynamics and Complexity*, 42–49. Springer.

Kleene, Stephen C. 1936. “General Recursive Functions of Natural Numbers.” In *The Undecidable*, edited by Martin Davis, 236–53. Raven Press.

———. 1939. “Review: The Logical Syntax of Language By Rudolf Carnap.” *The Journal of*

Symbolic Logic 4 (2): 82–87.

———. 1943. “Recursive Predicates and Quantifiers.” In *The Undecidable*, edited by Martin Davis, 254–87. Raven Press.

———. 1952. *Introduction to Metamathematics*. North-Holland.

———. 1967. *Mathematical Logic*. John Wiley & Sons.

———. 1981a. “Origins of Recursive Function Theory.” *Annals of the History of Computing* 3 (1): 52–67.

———. 1981b. “The Theory of Recursive Functions, Approaching Its Centennial.” *The Bulletin of the American Mathematical Society* 5 (1): 43–61.

———. 1987. “Reflections on Church’s Thesis.” *Notre Dame Journal of Formal Logic* 28 (4): 490–98.

———. 1988. “Turing’s Analysis of Computability, and Major Applications of It.” In *The Universal Turing Machine a Half-Century Survey*, edited by R. Herken, 17–54. Oxford University Press.

Kleene, Stephen C., and John B. Rosser. 1935. “The Inconsistency of Certain Formal Logics.” *The Annals of Mathematics* 36 (3): 630–36.

Koellner, Peter. 2006. “On the Question of Absolute Undecidability.” *Philosophia Mathematica* 14 (2): 153–88.

———. 2009a. “Carnap on the Foundations of Logic and Mathematics.” http://logic.harvard.edu/koellner/CFLM.pdf.

———. 2009b. “Truth in Mathematics: The Question of Pluralism.” In *New Waves in Philosophy of Mathematics*, edited by Otávio Bueno and Øystein Linnebo, 80–116. London: Palgrave Macmillan.

Kreisel, Georg. 1958. “Mathematical Significance of Consistency Proofs.” *The Journal of Symbolic Logic* 23 (2): 155–82.

———. 1965. “Mathematical Logic.” In *Lectures on Modern Mathematics*, edited by T. Saaty, 95–195. Wiley.

———. 1987. “Gödel’s Excursions into Intuitionistic Logic.” In *Gödel Remembered*, edited by P. Weingartner, 65–186. Bibliopolis.

Kremer, Michael. 1994. “The Argument of ‘On Denoting.’” *The Philosophical Review* 103 (2): 249–97.

Kripke, Saul A. 2013. “The Church-Turing ‘Thesis’ as a Special Corollary of Gödel’s Completeness Theorem.” In *Computability: Turing, Gödel, Church, and Beyond*, 77–104.

Löb, Martin H. 1955. “Solution of a Problem of Leon Henkin.” *The Journal of Symbolic Logic* 20 (2): 115–18.

Maddy, Penelope. 1980. “Perception and Mathematical Intuition.” *The Philosophical Review* 89 (2): 163–96.

———. 1990. *Realism in Mathematics*. Oxford University Press.

Makin, Gideon. 1995. “Making Sense of ‘On Denoting.’” *Synthese* 102: 383–412.

Mancosu, Paolo. 1998. “Hilbert and Bernays on Metamathematics.” In *The Adventure of Reason*, 125–58.

———. 2004. “Book Review: Kurt Gödel. Collected Works , Volumes IV and V.” *Notre Dame Journal of Formal Logic* 45 (2): 109–25.

Martin-Löf, Per. 2008. “The Hilbert-Brouwer Controversy Resolved?” In *One Hundred Years of Intuitionism (1907–2007)*, 243–56. North Holland.

Martin, Donald. 2005. “Gödel’s Conceptual Realism.” *The Bulletin of Symbolic Logic* 11 (2): 207–24.

Matiyasevich, Yuri. 1993. *Hilbert’s 10th Problem*. MIT Press.

Mendelson, Elliott. 1990. “Second Thoughts about Church’s Thesis and Mathematical Proofs.” *The Journal of Philosophy* 87 (5): 225–33.

Moore, Cristopher. 1990. “Unpredictability and Undecidability in Dynamical Systems.” *Physical Review Letters* 64 (20): 2354–57.

———. 1991. “Generalized Shifts: Unpredictability and Undecidability in Dynamical Systems.” *Nonlinearity* 4 (2): 199–230.

Moore, Gregory H. 1990. “Introductory Note to Gödel 1947 and 1964.” In *Kurt Gödel :*

Collected Works, Vol. II, 154–75.

Mostowski, A. 1966. *Thirty Years of Foundational Studies*. Oxford: Basil Blackwell.

Myhill, John. 1974. “The Undefinability of the Set of Natural Numbers in the Ramified Principia.” In *Bertrand Russell’s Philosophy*, edited by George Nakhnikian, 19–27. London: Duckworth.

Nagel, Ernest, and James R. Newman. 1958. *Gödel’s Proof*. New York: New York University Press.

Neale, Stephen. 1995. “The Philosophical Significance of Gödel’s Slingshot.” *Mind* 104 (416): 761–825.

Neumann, John von. 1927. “Zur Hilbertschen Beweistheorie.” *Mathematische Zeitschrift* 26 (1): 1–46.

Parsons, Charles. 1979. “Mathematical Intuition.” *Proceedings of the Aristotelian Society* 80. The Aristotelian Society: 145–68. doi:10.2307/4544956.

———. 1980. “Mathematical Intuition.” *Proceedings of the Aristotelian Society* 80. The Aristotelian Society: 145–68.

———. 1990. “Introductory Note to 1944.” In *Kurt Gödel : Collected Works, Vol. II*, 103–18.

———. 1992. “The Impredicativity of Induction.” In *Proof, Logic, and Formalization*, edited by Michael Detlefsen, 139–61.

———. 1995. “Platonism and Mathematical Intuition in Kurt Gödel’s Thought.” *The Bulletin of Symbolic Logic* 1 (1): 44–74.

———. 1998. “Finitism and Intuitive Knowledge.” In *The Philosophy of Mathematics Today*, edited by Mattias Schirn, 249–70. Oxford: Clarendon Press.

Péter, Rózsa. 1936. “Über Die Mehrfache Rekursion.” *Mathematische Annalen* 113 (1): 489–527.

Petzold, Charles. 2008. *The Annotated Turing : A Guided Tour Through Alan Turing’s Historic Paper on Computability and the Turing Machine*. Wiley.

Piccinini, Gualtiero. 2003. “Alan Turing and the Mathematical Objection.” *Minds and Machines* 13 (1): 23–48.

Pitowsky, Itamar. 1996. “Laplace’s Demon Consults an Oracle: The Computational Complexity of Prediction.” *Studies In History and Philosophy of Science Part B:* 27 (2): 161–80.

Poincaré, Henri. 1906. “Mathematics and Logic: III.” In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics, Vol. 2*, 1052–70.

———. 1910. “On Transfinite Numbers.” In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics, Vol. 2*, 1071–74.

Post, Emil. 1936. “Finite Combinatory Processes. Formulation 1.” In *The Undecidable*, edited by Martin Davis, 288–91. Raven Press.

———. 1944. “Recursively Enumerable Sets of Positive Integers and Their Decision Problems.” In *The Undecidable*, edited by Martin Davis, 304–37. Raven Press.

———. 1947. “Recursive Unsolvability of a Problem of Thue.” In *The Undecidable*, edited by Martin Davis, 292–303. Raven Press.

Potter, Michael. 2000. *Reason’s Nearest Kin*. Oxford University Press.

———. 2001. “Was Gödel a Gödelian Platonist?” *Philosophia Mathematica* 9 (3): 331–46.

Priest, Graham. 1994. “The Structure of the Paradoxes of Self-Reference.” *Mind* 103 (409): 25–34.

Putnam, Hilary. 1965. “Craig’s Theorem.” *The Journal of Philosophy* 62 (10): 251–60.

———. 1971. *Philosophy of Logic*. New York: Harper.

Quine, Willard V. 1936. “Truth by Convention.” In *Philosophical Essays for Alfred North Whitehead*, 90–124. New York: Longmans, Green and Co.

———. 1937. “New Foundations for Mathematical Logic.” *The American Mathematical Monthly* 44 (2): 70–80.

———. 1941. “Whitehead and the Rise of Modern Logic.” In *The Philosophy of Alfred North Whitehead*, edited by Paul Arthur Schilpp, 127–63.

———. 1951a. “Main Trends in Recent Philosophy: Two Dogmas of Empiricism.” *The Philosophical Review* 60 (1): 20–43.

———. 1951b. “On Carnap’s Views on Ontology.” *Philosophical Studies* 2 (5): 65–72.

———. 1955. “On Frege’s Way Out.” *Mind* 64 (254): 145–59.

———. 1960. “Carnap and Logical Truth.” *Synthese* 12 (4): 350–74.

———. 1966. “Russell’s Ontological Development.” *The Journal of Philosophy* 63 (21): 657–67.

———. 1971. “Epistemology Naturalized.” In *Akten Des XIV. Internationalen Kongresses Für Philosophie*, 6:87–103. Herder & Co.

Ramsey, Frank P. 1925. “The Foundations of Mathematics.” In *Foundations of Mathematics and Other Logical Essays*, 1–61. Routledge & Kegan Paul.

———. 1926. “Mathematical Logic.” *The Mathematical Gazette* 13 (184): 185–94.

———. 1929. “Philosophy.” In *Foundations of Mathematics and Other Logical Essays*, edited by Richard B. Braithwaite, 263–69. Routledge & Kegan Paul.

Rang, B., and W. Thomas. 1981. “Zermelo’s Discovery of the ‘Russell Paradox.’” *Historia Mathematica* 8 (1): 15–22.

Reid, Constance. 1970. *Hilbert*. Copernicus.

Rodriguez-Consuegra, Francisco, ed. 1995. *Kurt Gödel: Unpublished Philosophical Essays*. Springer Science & Business Media.

Rosenbloom, Paul. 1950. *The Elements of Mathematical Logic*. New York: Dover Publications.

Russell, Bertrand. 1903. *The Principles of Mathematics*. Cambridge University Press.

———. 1905. “On Denoting.” In *Essays in Analysis*, 103–19.

———. 1906a. “On ‘Insolubilia’ and Their Solution by Symbolic Logic.” In *Essays in Analysis*, 190–214.

———. 1906b. “On Some Difficulties in the Theory of Transfinite Numbers and Order Types.” In *Essays in Analysis*, 135–64.

———. 1907. “The Regresstive Method of Discovering the Premises of Mathematics.” In *Essays in Analysis*, 272–83.

———. 1908. “Mathematical Logic as Based on the Theory of Types.” In *From Frege to Gödel: A Source Book in Mathematical Logic*, edited by Jean van Heijenoort, 150–82.

———. 1910. “The Theory of Logical Types.” In *Essays in Analysis*, 215–54.

———. 1913. “The Philosophical Implications of Mathematical Logic.” In *Essays in Analysis*, 284–94.

———. 1919. *Introduction to Mathematical Philosophy*. London: George Allen & Unwin.

———. 1968. *The Autobiography of Bertrand Russell, 1914-1944*. London: Allen & Unwin.

———. 1994. *The Collected Papers of Bertrand Russell, vol.4 Foundations of Logic: 1903-1905*. Edited by Alasdair Urquhart. London and New York: Routledge.

Schilpp, Paul Arthur, ed. 1944. *The Philosophy of Bertrand Russell*. Open Court.

Searle, John R. 1958. “Russell’s Objections to Frege’s Theory of Sense and Reference.” *Analysis* 18 (6): 137–43.

Shagrir, Oron. 2002. “Effective Computation by Humans and Machines.” *Minds and Machines* 12 (2): 221–40.

———. 2006. “Gödel on Turing on Computability.” In *Church’s Thesis after 70 Years*, 393–419.

Shapiro, Stewart. 2013. “The Open Texture of Computability.” In *Computability: Turing, Gödel, Church, and Beyond*, 153–82.

Shoenfield, Joseph R. 1967. *Mathematical Logic*. Reading: Addison-Wesley.

Sieg, Wilfried. 1994. “Mechanical Procedures and Mathematical Experience.” In *Mathematics and Mind*, 71–117.

———. 1997. “Step by Recursive Step: Church’s Analysis of Effective Calculability.” *The Bulletin of Symbolic Logic* 3 (2): 154–80.

———. 1999. “Hilbert’s Programs: 1917–1922.” *The Bulletin of Symbolic Logic* 5 (1): 1–44.

———. 2002a. “Calculations by Man and Machine: Conceptual Analysis.” In *Reflections on the Foundations of Mathematics: Essays in Houor of Solomn Feferman*, 390–409.

———. 2002b. “Calculations by Man and Machine: Mathematical Analysis.” In *In the Scope of Logic, Methodology and Philosophy of Science, Vol. I*, 247–62.

———. 2005. “Only Two Letters: The Correspondence Between Herbrand and Gödel.” *The Bulletin of Symbolic Logic* 11 (2): 172–84.

———. 2006. “Gödel on Computability.” *Philosophia Mathematica* 14 (2): 189–207.

———. 2008. “Church Without Dogma: Axioms for Computability.” In *New Computational Paradigms*, edited by Barry Cooper, 139–52. New York: Springer.

———. 2009. “On Computability.” In *Philosophy of Mathematics. Handbook of the Philosophy of Science*, 535–630. Amsterdam: North-Holland.

———. 2013. “Gödel’s Philosophical Challenge (to Turing).” In *Computability: Turing, Gödel, Church, and Beyond*, 183–202.

Smith, Peter. 2013. *An Introduction to Gödel’s Theorems*. 2nd ed. Cambridge University Press.

Smoryński, Craig. 1988. “Hilbert’s Programme.” *CWI Quarterly* 1 (4): 3–59.

Soare, Robert I. 1996. “Computability and Recursion.” *The Bulletin of Symbolic Logic* 2 (3): 284–321.

———. 1999. “The History and Concept of Computability.” In *Handbook of Computability Theory 140*, 3–36.

Spector, Clifford. 1962. “Provably Recursive Functionals of Analysis: A Consistency Proof of Analysis by an Extension of Principles Formulated in Current Intuitionistic Mathematics.” *Recursive Function Theory*, 1–27.

Stein, Howard. 1988. “Logos, Logic, and Logistiké: Some Philosophical Remarks on Nineteenth-Century Transformation of Mathematics.” In *History and Philosophy of Modern Mathematics*, edited by William Aspray and Philip Kitcher, 238–59.

———. 1992. “Was Carnap Entirely Wrong, after All?” *Synthese* 93: 275–95.

Stoutland, Frederick. 2003. “What Philosophers Should Know about Truth and the Slingshot.” In *Realism in Action. Essays in the Philosophy of the Social Sciences*, 3–32. Dordrecht: Kluwer Academic Publishers.

Tait, William W. 1981. “Finitism.” *The Journal of Philosophy* 78 (10): 525–46.

———. 2001. “Gödel’s Unpublished Papers on Foundations of Mathematics.” *Philosophia Mathematica* 3 (9): 87–126.

———. 2006. “Gödel’s Correspondence on Proof Theory and Constructive Mathematics: Kurt Gödel. Collected Works. Volume IV: Selected Correspondence A-G; Volume V: Selected Correspondence H-Z.” *Philosophia Mathematica* 14 (1): 76–111.

———. 2010. “Gödel on Intuition and on Hilbert’s Finitism.” In *Kurt Gödel: Essays for His Centennial*, edited by Solomon Ferferman, 88–109. Cambridge University Press.

Takeuti, Gaisi. 2003. *Memoirs of a Proof Theorist: Gödel and Other Logicians*. Singapore: World Scientific.

Tarski, Alfred. 1936. “On the Concept of Logical Consequence.” In *Philosophy of Logic: An Anthology*, edited by Dale Jacquette, 210–16. Blackwell.

Tieszen, Richard. 1984. “Mathematical Intuition and Husserl’s Phenomenology.” *Noûs* 18 (3). Wiley: 395.

———. 2002. “Gödel and the Intuition of Concepts.” *Synthese* 133 (3): 363–91.

———. 2011. *After Gödel : Platonism and Rationalism in Mathematics and Logic*. Oxford University Press.

Turing, Alan. 1936. “On Computable Numbers, with an Application to the Entscheidungsproblem.” In *The Undecidable*, edited by Martin Davis, 115–54.

———. 1937. “Computability and λ-Definability.” *The Journal of Symbolic Logic* 2 (4): 153–63.

———. 1939. “Systems of Logic Based on Ordinals.” In *The Undecidable*, edited by Martin Davis, 154–222. Raven Press.

———. 1948. “Intelligent Machinery.” In *The Essential Turing*, edited by Jack Copeland, 395–432. Oxford University Press.

Uebel, Thomas. 2009. “Carnap’s Logical Syntax in the Context of the Vienna Circle.” In *Carnap’s Logical Syntax of Language*, edited by Pierre Wagner, 53–78. Palgrave Macmillan.

Urquhart, Alasdair. 1988. “Russell’s Zigzag Path to the Ramified Theory of Types.” *Russell: The Journal of Bertrand Russell Studies* 8 (1): 82–91.

Virdi, Arhat. 2009. “The Slingshot Argument, Gödel’s Hesitation and Tarskian Semantics.” *Prolegomena* 8 (2): 233–41.

Wang, Hao. 1974. *From Mathematics to Philosophy*. Routledge & Kegan Paul.

———. 1981. “Some Facts about Kurt Gödel.” *The Journal of Symbolic Logic* 46 (3): 653–59.

———. 1987. *Reflections on Kurt Gödel*. MIT Press.

———. 1996. *A Logical Journey: From Gödel to Philosophy*. MIT Press.

Webb, Judson. 1980. *Mechanism, Mentalism and Metamathematics : An Essay on Finitism*. Dordrecht: D. Reidel Publishing Company.

———. 1990. “Introductory Note to Remark 3 of Gödel (1972).” In *Kurt Gödel : Collected Works, Vol. II*, 292–304.

———. 2005. “Gödel’s Encounters with Formalism, Intuition, and Kant.” *Revue Internationale de Philosophie* 59 (234): 491–512.

Weyl, Hermann. 1921. “On the New Foundatioal Crisis of Mathematics.” In *From Brouwer To Hilbert: The Debate on the Foundations of Mathematics in the 1920s*, edited by Paolo Mancosu, 86–118. Oxford University Press.

———. 1925. “The Current Epistemological Situation in Mathematics.” In *From Brouwer To Hilbert: The Debate on the Foundations of Mathematics in the 1920s*, edited by Paolo Mancosu, 123–42. Oxford University Press.

———. 1946a. “Mathematics and Logic.” *The American Mathematical Monthly* 53 (1): 2–13.

———. 1946b. “The Philosophy of Bertand Russell. by P. A. Schilpp.” *The American Matheamtical Monthly* 53 (4): 208–14.

White, Morton, and Cordially Alfred. 1987. “A Philosophical Letter of Alfred Tarski.” *The Journal of Philosophy* 84 (1): 28–32.

Whitehead, Alfred North, and Bertrand Russell. 1927. *Principia Mathematica, Volume I*. Second Edi. Cambridge University Press.

Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus*. London: Kegan Paul.

Xing, Taotao. 2011. “How Gödel Relates Platonism to Mathematics.” *Theology and Science* 9 (1): 121–35.

Zach, Richard. 1998. “Numbers and Functions in Hilbert’s Finitism.” *Taiwanese Journal for Philosophy and History of Science* 10: 33–60.

———. 2001. “Hilbert’s Finitism: Historical, Philosophical and Metamathematical Perspectives.” University of California, Berkeley.

———. 2003. “Hilbert’s Program.” *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/hilbert-program/.

———. 2006. “Hilbert’s Program Then and Now.” In *Philosophy of Logic*, edited by Dov Gabbay and John Woods, 5:411–47. Elsevier.

Zermelo, Ernst. 1908. “Investigations in the Foundations of Set Theory I.” In *From Frege to Gödel: A Source Book in Mathematical Logic*, edited by Jean van Heijenoort, 199–215. Harvard University Press.

———. 1930. “On Boundary Numbers and Domains of Sets: New Investigations in the Foundations of Set Theory.” In *From Kant to Hilbert: A Source Book in the Foundations of Mathematics, Vol. 2*, edited by Bragg Ewald, 1219–33. Oxford University Press.